REMARKS

Claims 27-44 are pending in the application.

Reconsideration of the application is respectfully requested in view of the above amendments and the following remarks. For the Examiner's convenience, Applicants' remarks are presented in the order in which they were raised in the Office Action. Pertinent opinions of experts in the forms of a Second Declaration under 37 C.F.R. 1.132 from Dr. Amy Weiner, and a Declaration under 37 C.F.R. 1.132 from Dr. J.-H. James Ou, accompany this response.

A. Substance of Interview with the Examiner

Applicants thank Examiner Prouty for the interview granted on February 3, 2006 with SPE Kathleen Kerr, Examiner William Moore, Applicants' representative Gladys Monroy, and experts Dr. Amy Weiner and Dr. James Ou. An interview summary was mailed by the Examiner on March 1, 2006.

All pending claims were discussed in view of the art disclosed by Eckart et al., Pallaoro, et al. and Thiebault et al. Two primary issues were discussed: (1) With respect to whether the specification provides written description and enablement (how to use) the serine protease activity of SEQ ID NO: 65, Applicants argued that the Specification describes viral fragments and viral polyprotein as substrates and the art teaches that several substrates can be cleaved without the cofactor 4A. Agreement was not reached as the Examiner required that a suitable substrate for the serine protease according to one of skill in the art be addressed in the response. The Examiner's concerns are addressed in the following sections of this response.

(2) With respect to the issue of whether the specification provides sufficient description and enablement of an active NS2/3 protease as shown by Example 5 with reference to the SOD fusions p600, p500 and p300, Applicants argued that the protease activity shown in Example 5 of the specification corresponds to NS2/3 protease and addition of any additional sequence to the amino terminus at residue 946 (herein provided by hSOD) was sufficient to produce an active protease. Agreement was not reached as the Examiner required evidence of the stabilization of the

fusion constructs and explanation of the discrepancy in observed molecular weight be provided in the response. The Examiner's concerns are addressed in the following sections.

B. Non-Statutory Double Patenting

Claims 27-44 stand rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-14 of U.S. Patent No. 5,371,017.

Claim 36 remains provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claim 15 of copending Application No. 10/438,313, which is an application for reissue of U.S. Patent No. 5,371,017.

Applicants submit that they will file a terminal disclaimer in the present application to disclaim any term beyond the term of any earlier expiring patent in order to overcome this ground for rejection, after the conflicting claims are found to be allowable.

C. Claim Rejections under 35 U.S.C. § 112, First Paragraph

(i) Rejection of claims 27-44 for lack of Written Description under 35 U.S.C. § 112, first paragraph

Claims 27-44 remain rejected under 35 U.S.C. § 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, has possession of the claimed invention.

In particular, the Office Action states that the fusion proteins disclosed in Example 5 of Applicants' specification are insufficient for HCV-specific proteolysis even though they comprise specific amino acids His-952 and Cys-993 needed for NS2/NS3 protease cleavage. The Office Action raises the following arguments in support of its conclusion. Applicants' response to each of these arguments is also set forth below.

Applicants respectfully traverse. "[T]he 'essential goal' of the description of the invention requirement is to clearly convey the information that an Applicant has invented the subject matter which is claimed." *In re Barker*, 559 F.2d 588, 592 n.4, 194 USPQ 470, 473 n.4 (CCPA 1977). The test for sufficiency of support in a parent application is whether the disclosure of the application relied upon "reasonably conveys to the artisan that the inventor had possession at that time of the later claimed subject matter." *Ralston Purina Co. v. Far-Mar-Co., Inc.*, 772 F.2d 1570, 1575, 227 USPQ 177, 179 (Fed. Cir. 1985) (quoting *In re Kaslow*, 707 F.2d 1366, 1375, 217 USPQ 1089, 1096 (Fed. Cir. 1983)).

Applicants submit that the Specification of the '456 application discloses:

(i) a polynucleotide which encodes an NS3 domain hepatitis C virus protease that corresponds to the HCV NS3 serine protease activity, the protease comprising the entire domain required for NS3 serine protease, and discloses substrates for assaying its activity; and

(ii) a polynucleotide which encodes an NS3 domain hepatitis C virus protease that corresponds to the HCV NS2/3 protease activity, the protease comprising the active site residues of the NS2/3 protease and displaying an autocatalytic activity of the protein fused to a hSOD protein as disclosed in Example 5 of the Specification.

Either activity is sufficient for practicing a "method for assaying compounds for activity against hepatitis C virus."

While the Examiner appears to contend that because Applicants do not maintain that Example 5 of the Specification relates to a HCV NS3 serine protease activity, Applicants are required to rely solely on the NS2/3 protease activity in support of the claimed "NS3 domain HCV protease." Applicants respectfully disagree. Applicants contend that the "NS3 domain HCV protease" comprises both the NS2/3 protease activity disclosed in Example 5, and the NS3 serine protease activity disclosed *inter alia* in Examples 10 and 11.

(a) The Specification discloses an HCV NS3 domain that has a NS3 serine protease activity.

The Examiner concedes that the specification identifies and recognizes a serine protease activity. Office Action at pages 5-6. The Office Action notes that the specification defines an HCV NS3 domain protease as having termini established "***by expression and processing in an appropriate host*** of a DNA construct ***encoding the entire NS3 domain***." (emphasis original). Office Action at page 5.

Applicants submit that the '456 application not only discloses the structure of the NS3 serine protease, it also teaches a method for making it by *in vitro* expression. (Declaration of Dr. Ou, ¶¶ 9-11; Second Declaration of Dr. Weiner, ¶¶ 10-12; page 5, line 11 to page 6, line 4; page 7, line 19 through page 8, line 6; pages 8-9, 37-39, Tables 1 and 2).

Further, the Examiner contends that the publication of Eckart et al. does not support a constructive reduction to practice of a NS3 domain protease that either "comprises" or "consists of" SEQ ID NO:65 in the prophetic Examples 10 and 11 of the specification. According to the Office Action, the various HCV-derived proteins expressed in Eckart et al., whether or not the $Ser_{165}Gly$ mutation is present, share the entire NS3 domain and NS4A region, which are sufficient for cleaving at the NS3-NS4 boundary. Office Action, at 12-13. The Examiner suggests that the NS3 protease may not function in the absence of NS4A cofactor, which is absent from "the amino acid sequence of SEQ ID NO:65 or active NS3 domain hepatitis C virus protease truncation analog thereof." Applicants respectfully traverse.

While NS4A has been shown to act as a co-factor with NS3 serine protease, NS3 serine protease cleaves several substrates in the absence of the cofactor NS4A as discussed below. (Declaration of Dr. Ou, ¶¶ 13-18; Second Declaration of Dr. Weiner, ¶¶ 14-19.)

While it is now known that the NS3 serine protease cleaves the HCV polyprotein at multiple sites – NS3/4A, NS4A/4B, NS4B/5A and NS5A/5B, only the NS4B/5A cleavage is dependent on the presence of NS4A. Bartenschlager et al. (J Virol. 68(8):5045-5055 (1994)).

Bartenschlager has also shown that the first 211 amino acids of NS3 were sufficient for processing at all *trans* sites.

The NS3 serine protease-mediated cleavages at NS3/4A, NS4A/4B and NS5A/5B are processed efficiently in *trans* by the NS3 serine protease without NS4A.

> By using an NS3-5B substrate with an inactivated serine proteinase domain, trans-cleavage was observed at all sites except for the 3/4A site. Deletion of the inactive proteinase domain led to efficient trans-processing at the 3/4A site. Smaller NS4A-4B and NS5A-5B substrates were processed efficiently in trans; however, cleavage of an NS4B-5A substrate occurred only when the serine proteinase domain was coexpressed with NS4A.

Abstract, Lin et al., J. Virol. 68(12): 8147-8157 (1994).

Others have reported the same observation that NS4A is not a necessary co-factor for NS3 serine protease activity:

- Sardana et al. (Protein Expression and Purification 16:440-447 (1999)) showed that the NS4A cofactor is essential for "high" proteolytic activity of the NS3 serine protease. (see Abstract). However, Sardana also found that proteolysis at the NS4A/4B junction is carried out at detectable levels by the NS3 serine protease in the absence of NS4A. (Sardana at 443, left col.).

- Vishnuvardhan et al. (FEBS Lett. 402(2-3):209-212 (1997)) showed that a NS3 serine protease representing amino acids 1027–1218 of the HCV polyprotein, and not including any NS4A region, cleaves the NS5A/5B junction in the absence of NS4A. (Figs. 1 and 3). NS4A (amino acids 1658-1712; see Fig. 1) enhances the cleavage but is not essential for it. (Fig. 3). Further, Vishnuvardhan classifies the NS4A/4B cleavage site as "NS4A-independent" cleavage site. (at 211).

- Barbato et al. J. Mol. Biol. (1999) 289, 371-384, at 382, left col. states that "[i]nteraction with the NS4A cofactor is required to perform the cleavages at NS3/NS4A, NS4A/NS4B and NS4B/NS5A junctions but the proteinase in its

uncomplexed state is still able to cleave at the NS5A/NS5B boundaries, although with a much lower activity."

In their expert declarations, Drs. Ou and Weiner note that the functionally minimal domain required for activity of the NS3 serine protease is composed of 146 amino acids, residues 1059 to 1204 of the HCV polyprotein. (Yamada et al. Virology 246: 104-112 (1998)). Figure 1 and SEQ ID NO:1 (page 6, line 26 to page 7, line 18) of the '456 application discloses a sequence that encompasses the entire minimal domain of the NS3 serine protease. (Declaration of Dr. Ou, ¶ 18-20; Second Declaration of Dr. Weiner, ¶¶ 19-21.)

It has also been shown that the 146 amino acids long NS3 minimum domain can function by itself as a NS3 serine protease from a structural point of view. Love et al. Cell 87: 331-342 (1996).

Applicants respectfully submit that the specification discloses an "amino acid sequence of SEQ ID NO:65" and "NS3 domain hepatitis C virus protease or active NS3 domain hepatitis C virus protease truncation analog thereof" according to claims 1 and 6 that has HCV NS3 serine protease activity.

As discussed in detail below, under the enablement section, the specification also discloses substrates for "NS3 domain hepatitis C virus protease" such as full length viral polyprotein with the active residues disabled.

Since the specification discloses both a "NS3 domain hepatitis C virus protease" and a substrate therefor, Applicants submit that the specification indicates that the inventors were in possession of "an isolated polynucleotide which encodes an hepatitis C virus (HCV) proteolytic polypeptide, wherein said polypeptide consists essentially of an HCV NS3 domain protease or an HCV NS3 domain protease active truncation analog" and respectfully request withdrawal of this ground for rejection.

(b) The specification discloses an NS2/NS3 protease activity to those of ordinary skill in the art

The Examiner contends that the specification does not convey the existence of an NS2/NS3 metalloprotease to those of ordinary skill in the art. In particular, the Examiner states that the fusion proteins disclosed in Example 5 of specification are insufficient for HCV-specific proteolysis even though they comprise specific amino acids His-952 and Cys-993 needed for NS2/NS3 metalloprotease cleavage. Office Action at page 3.

It is uncontroverted that the fusion proteins on Example 1 contain 1-151 amino acids of human SOD protein and amino acids 946-1630 of the HCV polyprotein corresponding to the HCV NS3 domain protease sequence of Figure 1; the sequence corresponding to HCV includes C-terminal residues of NS2 and a majority of the NS3 sequence but not including the NS3/4 boundary; and the HCV sequence within this construct includes the putative NS2/3 cleavage site between Leu-1026 and Ala-1027 as well as the catalytic residues His-952 and Cys-993. Office Action at pages 3-4.

The Examiner does not agree that the specification inherently discloses a specific proteolytic activity that is native to the NS2/NS3 metalloprotease, and thus "the issue of whether the specification provides an adequate written disclosure of a claimed assay rests on whether the P600, P500, P300 and P190 proteins of Example 5 provide enough of the art-recognized structure of an HCV NS2/NS3 metalloprotease to permit cleavage at the art-recognized NS2/NS3 cleavage site that is present in each of these proteins." Office Action at pages 7-8. The Examiner argues that although certain NS2 sequences have been found to be necessary for NS2/NS3 metalloprotease activity (discussing the post-filing truncation studies of Hijikata et al., Grakoui et al., Reed et al., Santolini et al., Pieroni et al., Pallaoro et al. and Thibeault et al.), the fusion proteins of Example 5 lack these necessary NS2 sequences. Office Action at pages 8-11. The Office Action concludes that the specification fails to disclose or teach the structure of the portion of HCV polyprotein that is sufficient for NS2/NS3 metalloprotease-mediated cleavage at the NS2-NS3 boundary.

Applicants respectfully traverse. Applicants submit that Example 5 relates to the NS2/3 protease and not the NS3 serine protease. Applicants further submit that the observations of

Example 5 can be explained by the hSOD fusion portion of the Applicants' constructs being capable of replacing amino acids 898-946 of HCV NS2/3 to create an active protease.

From a review of the specification, one of skill in the art would understand that fusion of heterologous hSOD polypeptide sequence to a truncated NS2/3 protein, that by itself is inactive, restored activity of the NS2/3 protease activity, as discussed in the Declaration of Dr. Ou, ¶¶ 37-39; Second Declaration of Dr. Weiner, ¶¶ 37-39. Fusions with unrelated heterologous proteins are known to restore the activity of inactive proteins or stabilize truncated proteins:

- Fusion of a heterologous polypeptide sequence to a truncated fragment of a protein that by itself is inactive, can restore activity to the fusion protein fragment. A fragment containing the first domain of the CD45 protein lacks phosphatase activity, but fusion of this fragment to maltose-binding protein restores the phosphatase activity. Lorenzo et al., FEBS. 411(2-3):231-5 (1997).

- Fusion with hSOD had been observed to stabilize the HIV protease. (see Pichuantes et al. J. Biol. Chem 265(23), at p.13892, col. 2 (1990))

Since fusion of the NS2/3 fragments containing 299, 513 or 686 residues downstream from residue 946 to the C-terminal of a 151 amino acids long hSOD fragment displayed NS2/3 specific protease activity, as shown in Example 5, one of skill in the art would understand from Example 5 in the specification, that fusion of the heterologous hSOD sequence to the NS2/3 fragments containing the 299, 513 or 686 residues, was sufficient to restore NS2/3 specific protease activity. (Declaration of Dr. Ou, ¶¶ 39-40; Second Declaration of Dr. Weiner, ¶¶ 39-40.)

The crystal structure of the NS2/3 protease has revealed that it is a dimeric protein where each monomer has an amino-terminal subdomain containing two antiparallel alpha-helices (H1 and H2) and an active site is formed by a dimer interface comprising the critical residues for NS2-3 proteolytic activity, His-952 and Glu-972, located in the loop region following helix H2, and Cys-993 located in the b1–b2 loop of the C-terminal subdomain. (Lorenz et al., Nature 442:831-835, at para 3 of left column and para 1 of right column, and Figures 1 and 2, (2006); Declaration of Dr.

Ou, ¶41). The NS3 domain sequence of Figure 1 (SEQ ID NO: 70) includes all amino acids involved in dimerization and formation of the active site of the NS2/3 protease.

Applicants note that hSOD (M_r = 32,000) is a dimeric protein of 153 residues in each monomer. (Parge et al., Proc. Natl. Acad. Sci, USA 89:6109-6113 (1992).) The N-terminal 151 residues of hSOD are fused to the NS3 domain sequences in Examples 4 and 5 of the Specification.

Further, Applicants submit that stable and active viral protease fusion proteins were known in the art prior to 1991. For example, it was known in 1991 that fusion of heterologous sequences to the N-terminus of proteases does not affect the proteolytic activity of the protease. (Declaration of Dr. Ou, ¶¶ 34-35; Second Declaration of Dr. Weiner, ¶¶ 34-35). Human Immunodeficiency Virus (HIV) proteases remain active when a heterologous sequence is added to either terminus. The fused proteases mediate self-cleavage of viral polyproteins at the correct cleavage sites:

- A fusion protein comprising sequences from chloramphenicol acetyltransferase enzyme and HIV-1 protease is capable of autoprocessing, and mutation of the active site residue results in incorrect cleavage. Montgomery et al., Biochem. Biophys. Res. Comm., 175(3):784-94 (1991).

- An HIV protease fused to the amino or carboxy terminus of bacterial ß-galactosidase retains its capacity for specific autoprocessing. Valverde et al., J. Gen. Vir. 73:639-51 (1992)

(c) Cleavage mediated by the fusion proteins of Example 5 correspond to the HCV NS2/NS3 protease cleavage site

The Examiner contends that the calculated molecular mass of the observed cleavage product (24.9 kDa) "is clearly much less than the 34 kD relative molecular mass reported in the specification." Office Action, at page 14. Whereas the specification states that the size of this product is 34 kDa, the Examiner finds the molecular weight of the protein fragment corresponding to a predicted 232 amino acid cleavage product should be 24.9 kDa.

In response, Applicants submit that an anomaly in the estimates of molecular weights of proteins can be explained by a number of causes. Determination of exact molecular weight by SDS-polyacrylamide gel electrophoresis can be unpredictable. While SDS-polyacrylamide gel electrophoresis is often used to estimate molecular weights of proteins by comparing migration of proteins relative to a set of standard markers, it was well-known in 1991 that proteins and proteases do not necessarily migrate on SDS-polyacrylamide gels according to their predicted molecular weight. "[A]bnormalities in SDS binding or protein conformation, large differences in intrinsic protein charge, ... may lead to increased or decreased electrophoretic mobilities; therefore caution is advisable in use of this technique." Proteins: Structural and Molecular Principles. T. Creighton. page 33. (WH Freeman and Co., New York, © 1984). "[D]iscrepancy between apparent relative masses and real molecular weights underlies the uncertainty in deducing molecular masses of membrane-bound proteins from their mobility in electrophoretic gels." Introduction to Protein Structure. Brande C., and Tooze J. page 204 (Garland Publishing, Inc. New York and London © 1991). (Declaration of Dr. Ou, ¶¶ 29-30; Second Declaration of Dr. Weiner, ¶¶ 29-30).

In Example 5, the 34 kDa size is estimated from a Western blot of a SDS-polyacrylamide gel. ("[a] band corresponding to the hSOD fusion partner appeared at a relative molecular weight of about 34." the '456 application;" page 31, line 5 to page 32, line 12)

As discussed in the Declaration of Dr. Ou, ¶ 32; Second Declaration of Dr. Weiner, ¶ 31, several proteases, including a flavivirus NS2/3 protease, are known to migrate according to anomalous molecular weights in SDS-polyacrylamide gel electrophoresis:

- A NS2B-NS3 fusion protein from Dengue virus – a member of the flavivirus family which includes HCV – with a predicted molecular weight of 29.8 kDa displays anomalous migration in SDS-polyacrylamide gel electrophoresis with a higher apparent molecular mass of 37 kDa. Niyomrattanakit P., et al. J. Virol. (2004) 78(24): 13708-13716, at 13711, left column.

- A serine protease with a predicted molecular weight of 24.205 kDa was found to migrate at greater than 26 kDa possibly due to "the presence of bound [protein] defensin, possible posttranslational modifications of the protease, incomplete reduction of the protease during sample preparation or any combination of these possibilities." Hamilton JV et al., Insect Molecular Biology (2002) 11(3): 197–205, at 204, left column.

- The specification itself includes examples showing that estimates of molecular weights of known proteins from SDS-polyacrylamide gel electrophoresis were not precisely according to the predicted theoretical size. For example, the molecular weight of the 151 amino acid hSOD partner by itself was estimated by gel electrophoresis to be about 20 kDa at page 31, lines 15-1, whereas its theoretical size is 16.5 kDa.

One of skill in the art would have understood the "34 kDa" band to correspond to the product of specific cleavage by the NS2/3 protease from the consistent observation of a 34 kDa band reactive to anti-HCV antisera described in Example 5 of the specification of the '456 application, corresponding to the active fusion proteins P300, P500 and P600, but not with the inactive P190 fusion. (Declaration of Dr. Ou, ¶ 33; Second Declaration of Dr. Weiner, ¶ 33.)

One of skill in the art would have understood the inventors of the '456 application to have possession of both species -- a polynucleotide encoding the NS3 serine protease and a polynucleotide encoding the NS2/3 protease -- encompassed by a polynucleotide encoding the

"Hepatitis C Virus NS3 domain protease." Therefore, Applicants respectfully request withdrawal of this ground for rejection under 35 U.S.C. §112, ¶ 1, for lack of written description.

If the rejection is maintained, Applicants request the Examiner to provide an affidavit under 37 C.F.R. 1.104(d)(2) stating facts within the knowledge of the Examiner as to why the rejection should be maintained. Applicants reserve to right to explain or contradict the assertion with their own affidavits.

(ii) Rejection of claims 29, 30, 34, 35, 39 and 40 for lack of possession

Claims 29, 30, 34, 35, 39 and 40, drawn to genera of proteases comprising either SEQ ID NO:63 or SEQ ID NO:64, remain rejected as lacking written description for reasons of record. The Office Action contends that because the specification fails to disclose, suggest or teach any further features needed for the NS2/NS3 metalloprotease activity, the structural features of SEQ IDS NOS 63 and 64 (or the fusion proteins of Example 5 that comprise SEQ ID NOS: 63 and 64) are insufficient to produce a protein having the claimed proteolytic activity. Office Action, at page 15.

In response, Applicants submit that claims 29, 30, 34, 35, 39 and 40, drawn to a polynucleotide encoding a genera of proteases comprising either SEQ ID NO:63 or SEQ ID NO:64 are not directed to the NS2/3 protease activity, but instead relate to peptide sequences comprising essential active site residues for HCV NS3 serine protease activity. SEQ ID NO: 63 is an 11 amino acid sequence comprising the conserved HCV His-1083 residue of serine protease; and SEQ ID NO:64 is a 9 amino acid sequence comprising the conserved Ser-1165 residue of the HCV NS3 serine protease. These residues are essential for the catalytic activity of the serine protease. (the '456 application, page 7, line 19 through page 8, line 6 and Table 1). As discussed in detail above, the specification discloses an NS3 domain sequence comprising an active NS3 serine protease region and substrates there for. Claims 29, 30, 34 and 35 depend from claim 27, and claims 39 and 40 depend from claim 37 and further specify polynucleotides encoding the conserved regions comprising active site residues of the protease of claim 27 and 37. Therefore, Applicants respectfully request withdrawal of this ground for rejection.

(iii) Rejection of claims 27-44 for lack of enablement under 35 U.S.C. § 112

Claims 27-44 stand rejected under 35 U.S.C. § 112, first paragraph, because the specification allegedly fails to reasonably provide enablement for the preparation of polynucleotides of the claimed compositions or vectors encoding proteases that cleave at the NS2/3 cleavage site. According to the Office Action, the disclosed NS2 and NS3 domain regions present in the P600, P500, P300 and PI90 proteins that Applicant proposes are sufficient for the activity of an HCV NS2/NS3 metalloprotease have been shown by the discoveries of others to actually be insufficient for NS2/NS3 metalloprotease activity.

To be enabling, the specification of the patent must teach those skilled in the art how to make and use the full scope of the claimed invention without undue experimentation. (*Genentech Inc. v. NovoNordisk A/S* 108 F.3d 1361, 42 U.S.P.Q.2d 1001 (Fed. Cir. 1997)).

The Examiner states that the specification clearly does not teach how to make a polynucleotide encoding an HCV NS3 domain protease that has NS2/3 protease activity.

Applicants respectfully traverse. As discussed in detail in the previous section and in arguments filed in response to the first Office Action, Example 5 provides a method for making and using polynucleotides encoding the NS2/3 protease by fusion of a peptide having the sequence of Figure 1, or truncation analogs thereof, with a hSOD protein to demonstrate auto-catalytic activity shown in Example 5.

Further, Applicants note that as of the filing date of the parent application, April 4, 1991, vectors comprising polynucleotides that encode fusion of a protein of interest to human superoxide dismutase (hSOD) sequence was an established method of achieving high-level expression of a stable fusion protein. (Declaration of Dr. Ou, ¶¶ 35-36; Second Declaration of Dr. Weiner, ¶¶ 35-36). The specification of the '456 application discusses the expression from a vector of HIV protease as a fusion with human superoxide dismutase (hSOD) and having autocatalytic proteolysis activity by Pichuantes et al. (Specification, page 2, lines 13-20; Declaration of Dr. Ou, ¶ 35; Second Declaration of Dr. Weiner, ¶ 35).

Prior to 1991, examples of HIV proteases fused with hSOD and showing proteolytic activity for self-cleavage as well as cleavage using viral polyprotein substrates in *trans*, had been reported:

- hSOD-HIV2 protease fusion from bacteria and yeast correctly processes HIV-1 Pr53(gag) polyprotein in trans (Fig. 4). Pichuantes et al. J. Biol. Chem 265(23):13890-13898 (1990).

- A fusion protein of HIV1 protease with human superoxide dismutase (hSOD) expressed in yeast displayed correct self-processing, and trans-processing of gag-precursor Pr53gag substrate in *in vitro* assays (see Fig. 4, Table 1, Pichuantes et al., Proteins. 6:324-37 (1989)).

Applicants submit that the specification of the '456 application discloses a polynucleotide encoding a polypeptide sequence in Figure 1, which contains the active site residues of the NS2/3 protease, and the cleavage site of the NS2/3 protease. The specification also discloses how to make a NS2/3 protease by fusion with a hSOD partner which was a method that one of skill in the art in 1991 was familiar with. The hSOD fusion displayed autocatalytic cleavage corresponding to the expected NS2/3 cleavage site. Thus, the specification shows how to make and use a polynucleotide encoding an active HCV NS2/3 protease.

(b) Further, Applicants submit that the specification discloses a structure for a polynucleotide encoding the NS3 serine protease and how to find a useful substrate without undue experimentation.

The '456 application not only discloses the structure of a polynucleotide encoding the NS3 serine protease, it also teaches a method for using it for *in vitro* expression. (Declaration of Dr. Ou, ¶¶ 9-11; Second Declaration of Dr. Weiner, ¶¶ 10-12). Page 20, lines 14-16 of the specification discloses full-length polyprotein as a substrate for HCV protease. The specification also discloses use of alternative substrates in the form of "small peptides" (page 20, lines 21-26).

Applicants submit that viral polyprotein substrates for assaying proteases were commonly used in the art at the time of the invention. *See* Declaration of Dr. Ou, ¶ 21; Second Declaration of Dr. Weiner, ¶ 22. Protease assays using trans-cleavage of viral polyprotein substrates were known in the art at the time of the filing of the invention. Inactivation of the active site in the polyprotein substrate would enable one of skill in the art to assay protease activity in *trans*. (Declaration of Dr. Ou, ¶ 23; Second Declaration of Dr. Weiner, ¶ 24.)

The following examples show the widespread use of viral polyproteins as substrates for viral proteases prior to 1991:

- Processing of a 250 kDa Sindbis Virus polyprotein substrate (S1234) in vitro by Sindbis Virus protease prepared by in vitro translation. de Groot, et al. The EMBO J. 9(8)2631-2638 (1990):

- Trans-cleavage of a poliovirus capsomer precursor protein by poliovirus proteinase 3C. Nicklin: J. Virol (1988) 62: 4586-4593.

- Trans assay of MLV protease using Gazdar murine sarcoma virus (Gz-MSV) polyprotein Pr65(gag) substrate. Yoshinaka, Proc Natl Acad Sci U S A. (1985) 82(6):1618-1622.

- Trans assay of FeLV protease using Gazdar murine sarcoma virus (Gz-MSV) polyprotein Pr65(gag) substrate. Yoshinaka, J. Virol. (1985) 55(3):870-873.

- Trans assay of BLV protease using MLV polyprotein Pr65(gag) substrate. Yoshinoka *et al.*, J. Virol. (1986) 57(3):826-832.

- The proteinase of human immunodeficiency virus (HIV), expressed in Escherichia coli, shows rapid, efficient, and specific cleavage of an in vitro synthesized gag precursor polyprotein. Kräusslich et al., Proc Natl Acad Sci U S A. (1989) 86(3): 807–811.

- Processing of HIV-1 Pr53(gag) polyprotein substrate in trans (Fig. 4) by hSOD-HIV2 protease fusion from bacteria and yeast. Pichuantes et al. J. Biol. Chem 265(23):13890-13898 (1990)

- A fusion protein comprising HIV1 protease fused with human superoxide dismutase (hSOD) expressed in yeast displayed correct self-processing, and trans-processing in vitro of a gag-precursor Pr53gag polyprotein substrate. (see Fig. 4, Table 1, Pichuantes et al., Proteins. 6:324-37 (1989))

Applicants note that a substrate for the serine protease activity in the form of genomic HCV polyprotein is disclosed in page 20, lines 14-16 of the specification. Page 21, lines 4-5 explains that "[i]n the absence of this protease activity, the HCV polyprotein should remain in its unprocessed form." (Declaration of Dr. Ou, ¶ 22; Second Declaration of Dr. Weiner, ¶ 23.)

A method for inactivating the HCV protease activity in a HCV polyprotein by a single point mutation "substituting Ala for Ser121" is disclosed at page 22, line 27 to page 23, line 15 of the specification. One of skill in the art would have understood that this method can be used to inactivate the NS3 serine protease activity of the genomic HCV polyprotein – such that it can then be used as a substrate for testing NS3 serine protease activity in *trans*. (Declaration of Dr. Ou, ¶ 23; Second Declaration of Dr. Weiner, ¶ 24.)

In fact, the method disclosed in the specification was used by Lin et al. who used such a substrate with an inactivated serine proteinase domain to assay trans-cleavage by NS3 serine protease. (Lin et al., J. Virol. 68(12): 8147-8157 (1994)) (Declaration of Dr. Ou, ¶ 24; Second Declaration of Dr. Weiner, ¶ 25.)

Thus, one of skill in the art would understand that the '456 application describes a polynucleotide encoding a NS3 serine protease based on comparison with related flavivirus proteases and identification of critical amino acid residues of the serine triad. One of skill in the art would also understand that a substrate for the NS3 serine protease is disclosed in the '456

application in the form of genomic HCV polyprotein. (Declaration of Dr. Ou, ¶ 25; Second Declaration of Dr. Weiner, ¶ 26.)

Further, one of skill in the art would also understand that a substrate for the NS3 serine protease activity in the absence of NS4A cofactor is disclosed in the '456 application in the form of genomic HCV polyprotein. (Declaration of Dr. Ou, ¶ 25; Second Declaration of Dr. Weiner, ¶ 26.)

The '456 application teaches how to make and use a polynucleotide encoding both the NS3 serine protease and the NS2/3 protease encoded by the claimed Hepatitis C Virus NS3 domain protease or a truncation analog thereof, according to independent claims 27 and 37. Therefore, Applicants respectfully request withdrawal of this ground for rejection.

D. Claim Rejections under 35 U.S.C. § 112, Second Paragraph

(i) Rejection of claims 27-44 as being indefinite

According to the Office Action, claim 1 (*sic*) currently recites, "an NS3 domain hepatitis C virus *protease or an active . . . truncation analog*" (emphasis supplied in Office Action), but no sequence identifier is present in claim 1 to indicate the metes and bounds of the intended subject matter, such as where the truncation(s) might begin. See Office Action at pages 16-17. For instance, the public cannot determine what is more or less that the "domain" recited in the claims. Although Applicants argued that SEQ ID NO:70 should have the features sufficient for NS2/NS3 protease activity, the Office Action asserts that this sequence can alternatively, thus ambiguously, be construed to be a active truncation analog of an NS3 protease as well.

In response, Applicants submit that independent claims 27 and 37 specify a "HCV proteolytic polypeptide." Thus, one of skill in the art would understand that the "HCV NS3 domain protease truncation analog" need possess a "proteolytic" activity. The NS3 domain of HCV is characterized by the sequence of Figure 1 (SEQ ID NO: 70). The specification describes a truncation analog as: "the sequence may be substantially truncated, particularly at the carboxy terminus, apparently with full retention of protease activity." (page 8, lines 1-3).

As discussed in detail above, the proteins of Example 5 disclose a full length HCV domain protease (P600), truncation analogs that are active (P500, P300), and those that are not (P190). P600 includes all amino acids of Figure 1 (SEQ ID NO: 70). Thus, one of skill in the art would know that a truncation analog of SEQ ID NO: 70 would be a polypeptide that is missing amino acid residues from the full length NS3 domain but retains "proteolytic activity."

Figure 1 and SEQ ID NO:1 (pages 6-7) and SEQ ID NO:70 of the '456 application discloses a sequence that encompasses the entire minimal domain of the NS3 serine protease. (Declaration of Dr. Ou, ¶ 18-20; Second Declaration of Dr. Weiner, ¶¶ 19-21.) Further, SEQ ID NO: 65 which contains residues 1005 to 1204 of the HCV polyprotein, also includes all residues necessary for HCV NS3 serine protease activity.

In their expert declarations, Drs. Ou and Weiner note that the functionally minimal domain required for activity of the NS3 serine protease is composed of 146 amino acids, residues 1059 to 1204 of the HCV polyprotein. (Yamada et al. Virology 246: 104-112 (1998)). It has also been shown that the 146 amino acids long NS3 minimum domain can function by itself as a NS3 serine protease from a structural point of view. Love et al. Cell 87: 331-342 (1996). Thus, truncation analogs of the sequences of Fig. 1 and SEQ ID NO: 65 that retain NS3 serine protease activity would be readily available to one of skill in the art. Methods for truncation of an expressed amino acid sequence at either end by use of an expression vector and an exonuclease activity was routine in the art and disclosed at page 7, lines 28-31 of the Specification.

Therefore, Applicants respectfully request withdrawal of this ground for rejection under 35 U.S.C. § 112, Second Paragraph for alleged indefiniteness.

In addition, the Office Action raises a new ground of rejection of claims 27-31 as indefinite, allegedly necessitate by Applicants' amendment of claim 27 to recite the phrase "consisting essentially of." Office Action at page 17. The Office Action regards this term to be ambiguous and indefinite in the context of HCV protease. The Examiner states that a polynucleotide is a single covalently linked polymer and cannot consist essentially of more than one component.

"By using the term 'consisting essentially of,' the drafter signals that the invention necessarily includes the listed ingredients and is open to unlisted ingredients that do not materially affect the basic and novel properties of the invention." PPG Industries v. Guardian Industries, 156 F.3d 1351, 1354, 48 USPQ2d 1351, 1353-54 (Fed. Cir. 1998).

Applicants respectfully traverse. Applicants submit that claim 27 specifies a "polynucleotide which encodes an hepatitis C virus (HCV) proteolytic polypeptide." The polypeptide "consists essentially of" an "HCV NS3 domain protease ... ." The claim is directed to a polynucleotide that encodes a peptide sequence. A single polynucleotide can encode polypeptide sequences that have a "dominant" component ("HCV NS3 domain protease or an HCV NS3 domain protease active analog") and other amino acid sequences that do not materially affect the basic and novel properties of the invention may be encoded by the same polynucleotide. Therefore, Applicants respectfully request withdrawal of this ground for rejection.

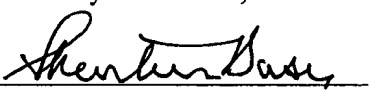
CONCLUSION

In view of the above, each of the presently pending claims in this application is believed to be in immediate condition for allowance. Accordingly, the Examiner is respectfully requested to withdraw the outstanding rejection of the claims and to allow this application to issuance. If it is determined that a telephone conference would expedite the prosecution of this application, the Examiner is invited to telephone the undersigned at the number given below.

In the event the U.S. Patent and Trademark Office determines that an extension and/or other relief is required, applicant petitions for any required relief including extensions of time and authorizes the Commissioner to charge the cost of such petitions and/or other fees due in connection with the filing of this document to Deposit Account No. 03-1952 referencing docket no. 223002010005. However, the Commissioner is not authorized to charge the cost of the issue fee to the Deposit Account.

Dated: February 12, 2007

Respectfully submitted,

By 

Shantanu Basu
Registration No.: 43,318
MORRISON & FOERSTER LLP
755 Page Mill Road
Palo Alto, California 94304-1018
(650) 813-5995

Atomic Structures of Wild-Type and Thermostable Mutant Recombinant Human Cu,Zn Superoxide Dismutase

HE Parge, RA Hallewell, and JA Tainer

PNAS 1992;89;6109-6113
doi:10.1073/pnas.89.13.6109

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6109–6113, July 1992
Biochemistry

Atomic structures of wild-type and thermostable mutant recombinant human Cu,Zn superoxide dismutase

(hydrogen bonds/protein design/helix dipole/metalloenzyme/protein conformation)

HANS E. PARGE*, ROBERT A. HALLEWELL†, AND JOHN A. TAINER*‡

*Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037; and †Chiron, 4560 Horton Street, Emeryville, CA 94608

Communicated by Jane S. Richardson, March 17, 1992 (received for review December 11, 1991)

ABSTRACT **Superoxide dismutase enzymes protect aerobic organisms from oxygen-mediated free-radical damage. Crystallographic structures of recombinant human Cu,Zn superoxide dismutase have been determined, refined, and analyzed at 2.5 Å resolution for wild-type and a designed thermostable double-mutant enzyme (Cys-6 → Ala, Cys-111 → Ser). The 10 subunits (five dimers) in the crystallographic asymmetric unit form an unusual stable open lattice with 80-Å-diameter channels. The 10 independently fit and refined subunits provide high accuracy, error analysis, and insights on loop conformations. There is a helix dipole interaction with the Zn site, and 14 residues form two or more structurally conserved side-chain to main-chain hydrogen bonds that appear critical to active-site architecture, loop conformation, and the increased stability resulting from the Cys-111 → Ser mutation.**

Superoxide is a normal by-product of aerobic metabolism and is produced in numerous reactions, including oxidative phosphorylation, photosynthesis, and the respiratory burst of stimulated neutrophils and macrophages (1, 2). The Cu,Zn superoxide dismutase (SOD) enzymes occur primarily in eukaryotes but have also been found in bacterial pathogens and symbionts. Major interests in SOD structure and function include the enzyme's Cu(II) and Zn(II) binding sites and active-site geometry (3), highly stable fold (4, 5), and extremely rapid reaction rate due to long-range electrostatic guidance of the anion substrate (6, 7). These essential SOD enzymes are critical components in the physiological response to oxygen toxicity and are actively investigated as potential therapeutic agents in pathological conditions related to oxidative stress [e.g., reperfusion damage following ischemia (8, 9), lung and tissue damage (10, 11), and general inflammation (12)]. Recently, effective mitigation of pulmonary oxygen toxicity resulted in Food and Drug Administration designation of SOD as an orphan drug for prevention of bronchopulmonary dysplasia in premature infants, and extensions to other applications are expected (13, 14). SOD is also involved in the pervasive bioregulatory functions of nitric oxide by preventing nitric oxide peroxidation by superoxide (15, 16). The biological and medical importance of SOD has stimulated efforts to develop mutant enzymes with potentially improved clinical effectiveness due to increased stability (17, 18) or serum half-life (19, 20). Identification of Cu,Zn SOD enzymes in parasites (21, 22) suggests the design of selective inhibitors that might block parasite SOD without affecting the human enzyme. All of these studies have been hampered by the lack of a structure for human SOD (HSOD), which has been cloned and expressed in yeast (23, 24). Here we report the determination, refinement, and analysis of the wild-type and the designed thermostable mutant human enzymes at 2.5 Å resolution.§

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Wild-type and thermostable mutant (Cys-6 → Ala, Cys-111 → Ser) HSODs were purified from recombinant yeast and crystallized in space group $C222_1$, $a = 205.2$, $b = 167.0$, $c = 145.5$ Å (25), with 10 subunits per asymmetric unit. Diffraction data comprising 742,579 observations of 77,344 unique reflections (88.4% complete to 2.48 Å) collected on the more stable mutant HSOD with a Xentronics area detector (Siemens Analytic X-Ray Instruments, Madison, WI) were used for the structure solution. The structure was solved by molecular replacement with the MERLOT (26) program package using the "humanized" dimer of bovine SOD as the probe. The solution was confirmed by using ethylmercuric phosphate, 3-chloromercuri-2-methylpropylurea, and potassium gold cyanide derivatives. The current model, which contains 1530 amino acids and 10 Cu, 10 Zn, and 2 sulfate ions and 499 water molecules, has been crystallographically refined with XPLOR (27) to a residual error (R factor) of 20.2% for 63,790 reflections ($>3\sigma$) between 10.0 and 2.5 Å resolution and 18.9% for 54,476 reflections ($>5\sigma$) between 6.0 and 2.5 Å resolution. Overall deviations from ideal geometry are 0.017 Å for bond distances and 3.5° for bond angles. For wild-type HSOD, 838,068 observations of 96,710 unique reflections were collected on film to 2.0 Å resolution using synchrotron radiation. This data is 81.5% complete (70,227 reflections) to 2.5 Å resolution for intensities $>2\sigma$. The refined model has an R factor of 21.0% for 69,714 reflections ($>3\sigma$) between 10.0 and 2.5 Å resolution, with overall deviations from ideal geometry of 0.023 Å for bond distances and 4.3° for bond angles.

RESULTS AND DISCUSSION

Crystal Lattice and Subunit Structure. HSOD ($M_r = 32,000$) is a dimeric enzyme with ellipsoidal dimensions of about 30 × 40 × 70 Å. Each identical subunit contains 153 residues, one Cu ion, and one Zn ion. The five HSOD dimers (named A, B, C, D, and E; each has subunits 1 and 2) pack to form a very open but well-ordered crystal lattice with over 68% solvent (Fig. 1). Each of the dimers is internally related by a noncrystallographic twofold axis approximately parallel to the crystallographic c axis. The five-dimer asymmetric units generate a honeycomb layer in the ab plane (Fig. 1*A*), composed of dog-bone-shaped building blocks consisting of two trimers of dimers, with the central C dimer participating in both trimers. These building blocks pack with two types of close interactions along whole subunit faces, at the ends and sides of the dog bone extremities (Fig. 1*B*). In the third dimension, along the c axis, the honeycomb layers are

Abbreviations: HSOD, human superoxide dismutase; SOD, superoxide dismutase.

‡To whom reprint requests should be addressed.

§The atomic coordinates for the 10 subunits of the thermostable mutant HSOD have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (file 1SOS).

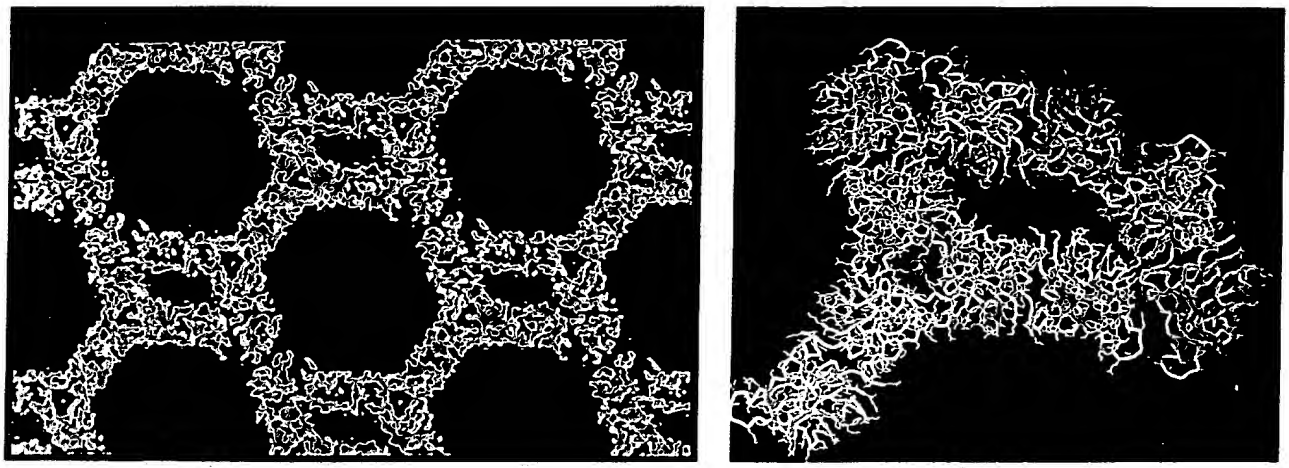


FIG. 1. Packing of the HSOD molecules in the crystal lattice, viewed down the *c* axis with the *a* axis horizontal and the *b* axis vertical. Five HSOD dimers (10 subunits) make up the crystallographic asymmetric unit. (*A*) Electron density map contours (blue) rendered in a solid raster view to show the interlocked packing of HSOD dimers in the *ab* plane, producing both high-resolution diffraction and the open lattice. The five-dimer, dogbone-shaped, asymmetric units (which lie along diagonals from the lower left to the upper right) are formed by two overlapping trimers of dimers. Packing of the next *ab* layer of the crystal lattice (not shown) constricts the open channels slightly, to about 80 Å in diameter. (*B*) The five HSOD dimers in an alternative asymmetric unit. The α-carbon backbones are shown with thicker tubes, and the individual side chains are shown with thinner tubes. Cu (gold sphere) and Zn (blue sphere) ions show the locations of the active sites. Subunits are color-coded in dimer pairs. The green A, blue B, and yellow C dimers form one trimer, with the A_2, B_2, and C_2 subunits associating around a sulfate group (red O and yellow S atoms). The purple D and red E dimers form two-thirds of the next trimer of dimers (upper right), completed by the C dimer of the adjacent asymmetric unit. The D_2, E_2, and C_1 subunits of this second trimer surround another sulfate group (upper right). The diamond-shaped links formed by the green, blue, purple, and red dimers form the zigzag chains running diagonally through the lattice from upper left to lower right in *A*. These chains are cross-braced by the yellow C dimers, which are exposed to the large lattice cavities and form the middle of the dog-bone building block. In the next layer of the crystal lattice, generated by the twofold screw axis along *c* through the blue B dimer, the next B dimer will pack on top of itself, but the next red E dimer will be offset below B and beside the yellow C dimer.

stacked on each other (by the twofold axis along *b*) and staggered by one dimer width along the *b* axis. This leaves only the E dimer without continuous contacts along *c* and reduces the cross-section of the hexagonal channels along *c* from 120 × 120 to 120 × 80 Å. There are other proteins that form open crystal lattices with large solvent channels (28, 29), but they are less highly ordered than the 2 Å resolution diffraction of these HSOD crystals.

The mutant and wild-type HSOD structures confirm that the overall flattened eight-stranded β-barrel fold, metal-site ligands and geometry, and dimer contact match that of the bovine SOD structure (30, 31). Superposition of the HSOD and bovine SOD structures indicated that the sequence changes do not alter the β-barrel diameter, strand angles, or loop conformations except in the region of the two-amino-acid insertion at sequence position 25. Loops form the active-site channel and connect the antiparallel β-strands with a Greek-key topology of +1, +1, +3, −1, −1, +3, +1.

Tenfold Redundancy. To assess structural accuracy and to distinguish conserved versus variable loop conformations in different crystal-packing environments, we did not average the electron density but instead fitted and refined each of the 10 HSOD subunits independently. This allows identification of specific patterns that provide the basis for loop conformation and stability, for this and other β-barrel proteins.

The unaveraged electron density shows structural details clearly, including carbonyl oxygens, bound water molecules, and crystal contact interactions (Fig. 2*A*). Similarly, the His-43 side-chain interactions that tie the short Greek-key connection to the active site (Fig. 2*B*) are well defined in the maps of each of the subunits. Superposition of the 10 subunits (Fig. 3*A*) shows that the loops and chain termini are only slightly more variable than the β-barrel framework. For all residues including crystal contacts, the average root-mean-square deviations in atomic positions are 0.38 Å for α-carbons, 0.43 Å for main-chain atoms, and 0.85 Å for all atoms. Atomic coordinates for all 10 subunits are available in the Protein Data Bank, for purposes of error analysis and for comparison with sets of multiple structures obtained by nuclear magnetic resonance methods.

Trp-32. The single tryptophan in HSOD is important in spectroscopic studies. It lies on the outside of the β-barrel, exposed in the dimer, but involved in some crystal contacts (Fig. 2*A*). Its conformation is completely invariant in all 10 subunits, however, with $\chi_1 \approx +60°$ and $\chi_2 \approx +90°$; one side of the ring lies against C^β and C^γ of Asn-19, while the other side and the ring N are exposed. In some but not all subunits, an ordered water hydrogen bonds with $N^{\epsilon 1}$. Nearby (at the bottom of Fig. 2*A*), there is a stripe of alternating charges: Glu-100, Lys-30, Glu-21, Lys-3, and the C terminus. Most of this environment around Trp-32 is conformationally invariant, but Lys-30 varies widely.

Critical Side-Chain Hydrogen Bonds. There are seven loops in SOD, numbered in sequence order. Loops I and V are short β-hairpin connections between adjacent β-strands. The five longer loops are shown in Fig. 3*B*. Loop II (residues 24–27) forms the β-hairpin containing the two-residue insertion relative to bovine SOD. Loops III (residues 37–40) and VI (residues 102–114) form the two Greek-key β-barrel connections. The active-site channel with its bound metal ions is formed between the electrostatic loop VII (121–144), implicated in substrate attraction, and loop IV (49–84), made up of the disulfide and the Zn-ligand subloop regions.

Compared to the β-barrel framework, the loops have more hydrogen bonds and other hydrophilic side-chain interactions. Most (>70 out of about 85) of the main-chain hydrogen bonds and ≥80 side-chain to main-chain hydrogen bonds in each subunit are conserved in all subunits, although most exposed polar side chains show significant conformational variability. At critical positions within or near the loops, 14 sequence-conserved (32), structurally conserved side chains appear to play important roles in loop conformation and interactions. Each forms two or more conserved side-chain to

FIG. 2. Atomic model (yellow bonds) for parts of the HSOD Cys-6 → Ala, Cys-111 → Ser thermostable double mutant, shown in typical electron density (pink and blue mesh). The residue numbering runs from 1 to 153 for subunit 1 and from 201 to 353 for subunit 2, preceded by the dimer letter. (*A*) Map and model showing β-strands 2–4 in subunit A_2 and their environment in the crystal. The three antiparallel β-strands run vertically, with clear electron density for side chains, carbonyl oxygens, and water molecules (indicated by +). The A_2–B_2 interface (upper left) includes a main-chain hydrogen bond between O of A_2 Asp-96 and N of B_2 Glu-132. The A_2–E_1 interface (right) involves several side-chain hydrogen bonds, including the $N^{\epsilon 1}$ of A_2 Trp-32 to $O^{\delta 2}$ of E_1 Asp-96. Although buried in this contact, Trp-32 (an important spectroscopic probe) is exposed on the surface of the dimer, as can be seen in Fig. 1*B*. (*B*) Map and model of the short Greek-key connection (loop III) and the adjacent active site, showing the key side-chain interactions of His-43 (labeled A243). The Cu(II) and Zn(II) ions are in high peaks of electron density (pink contours) above and below bridging His-63. Loop III curves around the His-43 ring at top, and the following β-strand continues down to Cu-ligand His-46. His-43 links the Greek-key connection to the β-barrel and the active site by hydrogen bonds from its ring nitrogens to the backbone carbonyls of Thr-39 and Cu-ligand His-120.

main-chain hydrogen bonds to nonadjacent residues or ligates a metal ion and forms one such hydrogen bond.

Six of these critical side chains form hydrogen bond bridges between structural elements (Fig. 3*B*). Five are on β-strands, and one is in a loop near the beginning of a strand (Arg-143); all anchor the active-site channel by hydrogen bonding to main-chain atoms, usually O, of the loops (Fig. 3*B*). His-43 is shown in detail in Fig. 2*B*. Cu-ligands His-48 and His-120 hydrogen bond to the disulfide subloop at O_{61} and to the electrostatic loop at O_{141} and O_{142}, respectively.

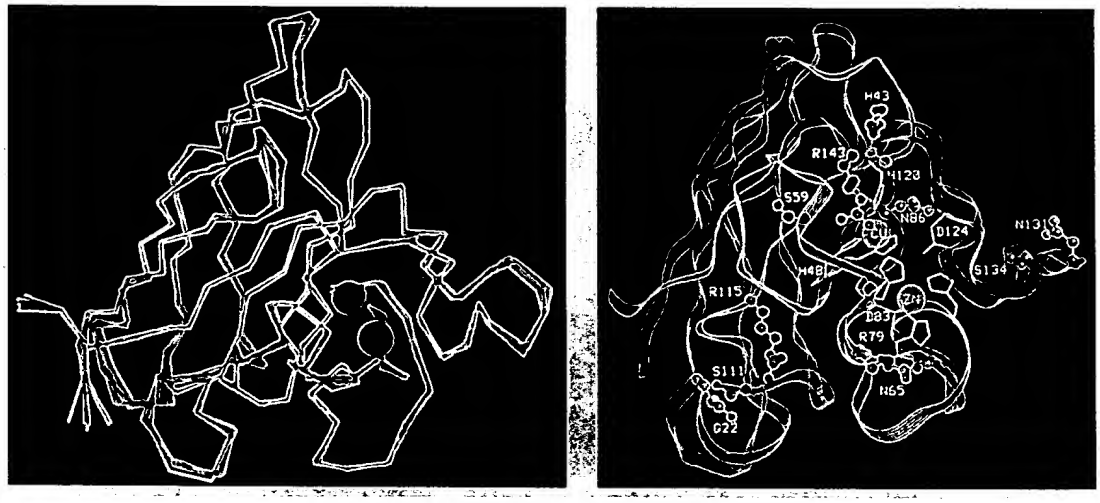


FIG. 3. Conservation and variation of the subunit fold and conformationally critical side chains shown for the thermostable double mutant of HSOD. (*A*) The α-carbon backbones for the 10 superimposed HSOD subunits, showing the overall conformational accuracy. This 10-fold redundancy provides confident assignment of specific loop conformations and major side-chain interactions and also gives a data set suitable for comparison with multiple structure sets obtained for other proteins by NMR methods. The greatest variations in HSOD occur at the loop bends and the two termini. In the active-site channel around the Cu (gold sphere) and Zn (blue sphere), the loops superimpose closely. Two turns of α-helix (horizontal, at far right) provide helix dipole stabilization of the Zn(II) ion. (*B*) HSOD β-structure framework (blue) and loop regions, along with conformationally important side chains. The α-carbon backbone is shown as a ribbon colored to highlight the various loops. Critical side chains that form multiple side-chain to main-chain hydrogen bonds are shown in ball and stick representation, labeled by residue number and the one-letter amino acid code and color-coded to match the loops they stabilize: Gln-22 for the purple insertion loop II (see Fig. 4*A*); His-43 for the light-blue Greek-key connection (loop III; see also Fig. 2*B*); Ser-59 and Arg-143 for the yellow disulfide subloop (IV); Asn-65 and Arg-79 for the gold Zn-ligand subloop (IV); Ser-111 and Arg-115 for the green Greek-key loop VI; and Asn-86, Asn-131, and Ser-134 for the red electrostatic loop VII. Metal ligands His-48, His-120, and Asp-83 also make side-chain to main-chain hydrogen bonds to the loop regions. The Cu-ligands (His-46, His-48, His-63, and His-120) form distorted square-planar geometry, while the Zn-ligands (His-63, His-71, His-80, and Asp-83) are tetrahedral. The Cu and Zn are linked directly by the bridging His-63 and indirectly by the side-chain carboxylate of buried Asp-124, which hydrogen bonds to both a Cu and a Zn-ligating histidine.

Asn-86 forms one hydrogen bond to the β-strand Gly-44 nitrogen and two to the main-chain atoms of electrostatic loop residue Asp-124, whose side chain in turn bridges between Zn-ligand His-71 and Cu-ligand His-46. Arg-115 hydrogen bonds to O_{111} in the longer Greek-key loop and to Cu-ligand neighbor O_{49}. Activity-important Arg-143 in the electrostatic loop hydrogen bonds to three backbone carbonyls in the disulfide subloop.

One other structurally conserved hydrogen bond links loops IV and VII to brace the active-site channel, from the backbone N of Zn-ligand His-71 to the main-chain CO of Thr-135 in the short helix in the electrostatic loop. This six-residue helix (residues 132–137) is well-ordered and has a consistent conformation in the 10 independent subunits. The α-helix dipole is oriented to stabilize and be stabilized by the Zn-ion binding site (Fig. 3*B*). This helix dipole interaction, which was not recognized in structural analyses of bovine SOD, suggests that Zn binding stabilizes the appropriate conformation for the electrostatically important, sequence-conserved residues Glu-132, Glu-133, and Lys-136 shown computationally (7, 33) to promote recognition of the superoxide anion substrate.

Eight of the sequence-conserved, structurally consistent side chains form multiple side-chain to main-chain hydrogen bonds across bends within a loop, presumably to control conformation for each of the longer loop regions (Fig. 3*B*). The Ser-59 O^{γ} hydrogen bonds to O_{52}, N_{54}, N_{55}, and O_{56} to cross-brace the major bend in the disulfide subloop. The Zn-ligand subloop contains hydrogen bonds from the Asn-65 side chain to N_{68}, N_{69}, and O_{69}; from the Arg-79 side chain to O_{74}, O_{81}, and O_{83}; and from the Asp-83 side chain to N_{72} and N_{80}. Within the electrostatic loop, the Asn-131 O^{δ} hydrogen bonds to N_{133} and N_{134} as the initiating N-cap of the helix discussed above. The side chain of Ser-134 stabilizes the major open turn within the electrostatic loop by hydrogen bonding to O_{127}, N_{128}, and N_{129}.

If side-chain to main-chain hydrogen bonds help control conformation at longer loop bends, then they might be expected at loop insertions. Indeed, the two-residue insertion (relative to bovine SOD) in loop II of HSOD is stabilized by hydrogen bonds from the side chain of Gln-22 across the β-hairpin to O_{25} and O_{27} (Fig. 4*A*). Gln-22 is present in all sequences containing the loop insertion, suggesting that it is a compensatory sequence change. Ala-22 → Gln is in fact the only nonconservative sequence difference between bovine SOD and HSOD on the first four β-strands that is not fully solvent-exposed. Such mechanisms to accommodate loop insertions may be needed in evolution, since, for instance, random single-residue loop insertions in staphylococcal nuclease are significantly destabilizing (with an average cost of 3.8 kcal/mol) (34).

Thermostability of the Double Mutant. To experimentally test the role of free disulfides in irreversible thermal inactivation and the role of intraloop hydrogen bonds in providing thermodynamic stabilization, we mutated both free cysteines in HSOD, separately and together, to residues naturally occurring in these positions in other SOD sequences: Cys-6 → Ala and Cys-111 → Ser. All three mutants are more stable to irreversible thermal inactivation, presumably due to the removal of the reactive thiols (18); the double mutant is more stable to irreversible thermal inactivation than either of the single mutants.

In wild-type HSOD, Cys-111 forms weak (long) hydrogen bonds to O_{106} and N_{113} in only 2 of the 10 subunits, whereas in the thermostable HSOD double mutant, Ser-111 consistently makes both of these hydrogen bonds that stabilize the Greek-key loop (Fig. 4*B*). Comparison of the wild-type and double mutant HSOD structures at position 6 shows shifts similar to those observed in the bovine SOD Cys-6 → Ala structure (17), which completely close the cavity previously occupied by the sulfur atom. This results in a slight increase (0.1 kcal/mol) in stability to reversible denaturation relative to wild-type HSOD (18). In contrast, the HSOD Cys-111 → Ser single mutant, which not only removes a reactive thiol but also provides improved side-chain to main-chain hydrogen bonding, is 0.8 kcal/mol more stable than the wild type (18). These structure and stability results suggest that side-chain to

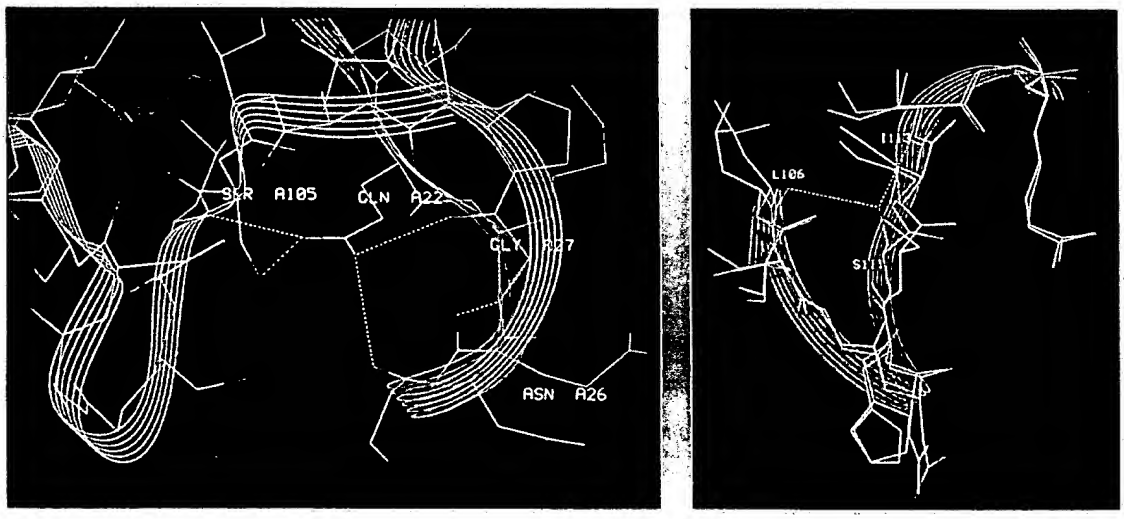


FIG. 4. Critical pairs of side-chain to main-chain hydrogen bonds stabilize the two-residue insertion in loop II and loop VI of the thermostable double mutant. The structural model is shown with bonds colored by atom type: red, oxygen; blue, nitrogen; and green, carbon. (*A*) The two-residue insertion of Asn-26 and Gly-27 (white ribbon) into loop II is stabilized by a set of hydrogen bonds from Gln-22 (center) that link the carbonyl oxygen atoms of loop residues Ser-25 and Gly-27 back to the nearby Greek-key loop at Ser-105 and Leu-106 (whose side chain plugs one end of the β-barrel). The SODs from other species that have this two-residue loop insertion all have Gln in position 22. (*B*) Superposition of the wild-type (pink) and thermostable double mutant (atom-colored) structures in Greek-key loop VI. The mutant has increased conformational stability resulting in part from stronger side-chain to main-chain hydrogen bonds from the O^{γ} of mutated Ser-111 to the N of Ile-113 and the O of Leu-106. Similar hydrogen bonds are formed in the bovine SOD structure (not shown). In wild-type HSOD (pink), Cys-111 and Leu-106 are pushed away from each other, and the Greek-key opens up. Although long hydrogen bonds are formed by Cys-111 in two of the 10 wild-type subunits, this loop is not consistently tethered in the wild type as it is in the mutant structure.

main-chain hydrogen bonds across loop bends can provide a net stabilization to the folded protein.

Implications for SOD Structure and Protein Loop Stability and Design. The activity and stability of Greek-key β-barrel proteins, such as HSOD, are strongly determined by the loops joining the β-strand framework. Loops have been categorized by structural patterns (35), and roles of individual residues in controlling antibody loop conformations have been identified (36, 37). The present results identify 14 sequence-conserved, structurally consistent residues that form multiple side-chain to main-chain hydrogen bonds critical for HSOD active-site stereochemistry, for local loop conformations, and for the stable connection of different structural elements. Reliable characterization of these critical side chains benefited greatly from their structural consistency in the 10 independent HSOD subunits. Comparison of the bovine SOD and HSOD structures suggests that side-chain to main-chain hydrogen bonds can compensate for loop insertions. The structure and stability of the (HSOD Cys-6 → Ala, Cys-111 → Ser) mutant demonstrate that intraloop side-chain to main-chain hydrogen bonds can thermodynamically stabilize the enzyme. The hydrogen bonding patterns and the α-helix dipole interaction tie the Zn site to the loop involved in electrostatic recognition of the anion substrate. Taken together, these results provide a structural basis for loop redesign in HSOD and other β-barrel proteins, in order to alter activity, stability, and binding functions.

We thank Drs. Elizabeth Getzoff, Leslie Kuhn, Duncan McRee, Victoria Roberts, and David Stout for discussion; Mark Yeager and John Berriman for electron microscopy to identify the threefold symmetry along the crystallographic *c* axis; Michael Pique for help with computer graphics; Paul Phizackerley, Michael Soltis, and Henry Bellamy at the Stanford Synchrotron Radiation Laboratory for support during data collection; and BIOSYM for the use of their program INSIGHT. This work was supported by National Institutes of Health Grant GM39345 (to J.A.T. and R.A.H.).

1. Fridovich, I. (1986) in *Advances in Enzymology*, ed. Meister, A. (Wiley, New York), Vol. 58, pp. 61–97.
2. Halliwell, B. & Gutteridge, J. M. C. (1989) *Free Radicals in Biology and Medicine* (Clarendon Press, Oxford).
3. Tainer, J. A., Roberts, V. A., Fisher, C. L., Hallewell, R. A. & Getzoff, E. D. (1991) in *A Study of Enzymes*, ed. Ruby, S. A. (CRC Press, Boca Raton), pp. 499–538.
4. Roe, J. A., Butler, A., Scholler, D. M., Valentine, J. S., Marky, L. & Breslauer, K. J. (1988) *Biochemistry* **27**, 950–958.
5. Malinowski, D. P. & Fridovich, I. (1979) *Biochemistry* **18**, 5055–5060.
6. Cudd, A. & Fridovich, I. (1982) *J. Biol. Chem.* **257**, 11443–11447.
7. Getzoff, E. D., Tainer, J. A., Weiner, P. K., Kollman, P. A., Richardson, J. S. & Richardson, D. C. (1983) *Nature (London)* **306**, 287–290.
8. Murohara, Y., Yui, Y., Hattori, R. & Kawai, C. (1991) *Am. J. Cardiol.* **67**, 765–767.
9. Zweier, J. L., Flaherty, J. T. & Weisfeldt, M. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1404–1407.
10. White, C. W., Avraham, K. B., Shanley, P. F. & Groner, Y. (1991) *J. Clin. Invest.* **87**, 2162–2168.
11. Oda, T., Akaike, T., Hamamoto, T., Suzuki, F., Hirano, T. & Maeda, H. (1989) *Science* **244**, 974–976.
12. Halliwell, B., Gutteridge, J. M. C. & Blake, D. (1985) *Philos. Trans. R. Soc. London B* **311**, 659–671.
13. Walther, F. J., Kuipers, I. M., Pavlova, Z., Willebrand, D., Abuchowski, A. & Viau, A. T. (1990) *Exp. Lung Res.* **16**, 177–189.
14. Davis, J. M., Rosenfeld, W., Gonenne, A., Whitin, J., Becker, J., Metlay, L. & Penney, D. (1991) *Pediatric Res.* **29**, 313 (Abstr.).
15. Beckman, J. S. (1990) *Nature (London)* **345**, 27–28.
16. Beckman, J. S., Beckman, T. W., Chen, J., Marshall, P. A. & Freeman, B. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1620–1624.
17. McRee, D. E., Redford, S. M., Getzoff, E. D., Lepock, J. R., Hallewell, R. A. & Tainer, J. A. (1990) *J. Biol. Chem.* **265**, 14234–14241.
18. Lepock, J. R., Frey, H. E. & Hallewell, R. A. (1990) *J. Biol. Chem.* **265**, 21612–21618.
19. Ando, Y., Inoue, M., Araki, S. & Morino, Y. (1991) *Biochim. Biophys. Acta* **1073**, 374–379.
20. Hallewell, R. A., Laria, I., Tabrizi, A., Carlin, G., Getzoff, E. D., Tainer, J. A., Cousens, L. S. & Mullenbach, G. T. (1989) *J. Biol. Chem.* **264**, 5260–5268.
21. Henkle, K. J., Liebau, E., Müller, S., Bergmann, B. & Walter, R. D. (1991) *Infect. Immun.* **59**, 2063–2069.
22. Simurda, M. C., vanKeulen, H., Rekosh, D. M. & LoVerde, P. T. (1988) *Exp. Parasitol.* **67**, 73–84.
23. Hallewell, R. A., Mills, R., Tekamp-Olson, P., Blacher, R., Rosenberg, S., Ötting, F., Masiarz, F. R. & Scandella, C. J. (1987) *Biotechnology* **5**, 363–366.
24. Hallewell, R. A., Masiarz, F. R., Najarian, R. C., Puma, J. P., Quiroga, M. R., Randolph, A., Sanchez-Pescador, R., Scandella, C. J., Smith, B., Steimer, K. S. & Mullenbach, G. T. (1985) *Nucleic Acids Res.* **13**, 2017–2034.
25. Parge, H. E., Getzoff, E. D., Scandella, C. S., Hallewell, R. A. & Tainer, J. A. (1986) *J. Biol. Chem.* **261**, 16215–16218.
26. Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr.* **21**, 273–278.
27. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235**, 458–460.
28. White, S. P., Cohen, C. & Phillips, G. N., Jr. (1987) *Nature (London)* **325**, 826–828.
29. Carter, D. C., He, X.-M., Munson, S. H., Twigg, P. D., Gernert, K. M., Broom, M. B. & Miller, T. Y. (1989) *Science* **244**, 1195–1198.
30. Tainer, J. A., Getzoff, E. D., Richardson, J. S. & Richardson, D. C. (1983) *Nature (London)* **306**, 284–287.
31. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. (1982) *J. Mol. Biol.* **160**, 181–217.
32. Getzoff, E. D., Tainer, J. A., Stempien, M. M., Bell, G. I. & Hallewell, R. A. (1989) *Proteins: Struct. Funct. Genet.* **5**, 322–336.
33. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L. & Hallewell, R. A. (1992) *Nature (London)*, in press.
34. Sondek, J. & Shortle, D. (1990) *Proteins: Struct. Funct. Genet.* **7**, 299–305.
35. Leszczynski, J. F. & Rose, G. (1986) *Science* **234**, 849–855.
36. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989) *Nature (London)* **342**, 877–883.
37. Chien, N. C., Roberts, V. A., Guisti, A., Scharff, M. D. & Getzoff, E. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5532–5536.

Biochemistry. In the article "Atomic structures of wild-type and thermostable mutant recombinant human Cu,Zn superoxide dismutase" by Hans E. Parge, Robert A. Hallewell, and John A. Tainer, which appeared in number 13, July 1992, of *Proc. Natl. Acad. Sci. USA* (89, 6109–6113), the authors request that the following corrections be noted. Due to printer's errors, Fig. 1 was incorrectly cropped and Fig. 2 was not in focus. The corrected figures are shown below. In addition, in Figs. 1–4, locants *A* and *B* were inadvertently left off; therefore, throughout the paper, panels *A* and *B* in Figs. 1–4 correspond to the left and right panels, respectively.

FIG. 1. Packing of the HSOD molecules in the crystal lattice, viewed down the *c* axis with the *a* axis horizontal and the *b* axis vertical. Five HSOD dimers (10 subunits) make up the crystallographic asymmetric unit. (*A*) Electron density map contours (blue) rendered in a solid raster view to show the interlocked packing of HSOD dimers in the *ab* plane, producing both high-resolution diffraction and the open lattice. The five-dimer, dogbone-shaped, asymmetric units (which lie along diagonals from the lower left to the upper right) are formed by two overlapping trimers of dimers. Packing of the next *ab* layer of the crystal lattice (not shown) constricts the open channels slightly, to about 80 Å in diameter. (*B*) The five HSOD dimers in an alternative asymmetric unit. The α-carbon backbones are shown with thicker tubes, and the individual side chains are shown with thinner tubes. Cu (gold sphere) and Zn (blue sphere) ions show the locations of the active sites. Subunits are color-coded in dimer pairs. The green A, blue B, and yellow C dimers form one trimer, with the A_2, B_2, and C_2 subunits associating around a sulfate group (red O and yellow S atoms). The purple D and red E dimers form two-thirds of the next trimer of dimers (upper right), completed by the C dimer of the adjacent asymmetric unit. The D_2, E_2, and C_1 subunits of this second trimer surround another sulfate group (upper right). The diamond-shaped links formed by the green, blue, purple, and red dimers form the zigzag chains running diagonally through the lattice from upper left to lower right in *A*. These chains are cross-braced by the yellow C dimers, which are exposed to the large lattice cavities and form the middle of the dog-bone building block. In the next layer of the crystal lattice, generated by the twofold screw axis along *c* through the blue B dimer, the next B dimer will pack on top of itself, but the next red E dimer will be offset below B and beside the yellow C dimer.

FIG. 2. Atomic model (yellow bonds) for parts of the HSOD Cys-6 → Ala, Cys-111 → Ser thermostable double mutant, shown in typical electron density (pink and blue mesh). The residue numbering runs from 1 to 153 for subunit 1 and from 201 to 353 for subunit 2, preceded by the dimer letter. (*A*) Map and model showing β-strands 2–4 in subunit A_2 and their environment in the crystal. The three antiparallel β-strands run vertically, with clear electron density for side chains, carbonyl oxygens, and water molecules (indicated by +). The A_2–B_2 interface (upper left) includes a main-chain hydrogen bond between O of A_2 Asp-96 and N of B_2 Glu-132. The A_2–E_1 interface (right) involves several side-chain hydrogen bonds, including the $N^{\varepsilon 1}$ of A_2 Trp-32 to $O^{\delta 2}$ of E_1 Asp-96. Although buried in this contact, Trp-32 (an important spectroscopic probe) is exposed on the surface of the dimer, as can be seen in Fig. 1*B*. (*B*) Map and model of the short Greek-key connection (loop III) and the adjacent active site, showing the key side-chain interactions of His-43 (labeled A243). The Cu(II) and Zn(II) ions are in high peaks of electron density (pink contours) above and below bridging His-63. Loop III curves around the His-43 ring at top, and the following β-strand continues down to Cu-ligand His-46. His-43 links the Greek-key connection to the β-barrel and the active site by hydrogen bonds from its ring nitrogens to the backbone carbonyls of Thr-39 and Cu-ligand His-120.

Vol 442|17 August 2006|doi:10.1038/nature04975

nature

LETTERS

Structure of the catalytic domain of the hepatitis C virus NS2-3 protease

Ivo C. Lorenz[1]*, Joseph Marcotrigiano[1]*, Thomas G. Dentzer[1] & Charles M. Rice[1]

Hepatitis C virus is a major global health problem affecting an estimated 170 million people worldwide[1]. Chronic infection is common and can lead to cirrhosis and liver cancer. There is no vaccine available and current therapies have met with limited success[2]. The viral RNA genome encodes a polyprotein that includes two proteases essential for virus replication[3,4]. The NS2-3 protease mediates a single cleavage at the NS2/NS3 junction, whereas the NS3-4A protease cleaves at four downstream sites in the polyprotein. NS3-4A is characterized as a serine protease with a chymotrypsin-like fold[5,6], but the enzymatic mechanism of the NS2-3 protease remains unresolved[7–9]. Here we report the crystal structure of the catalytic domain of the NS2-3 protease at 2.3 Å resolution. The structure reveals a dimeric cysteine protease with two composite active sites. For each active site, the catalytic histidine and glutamate residues are contributed by one monomer, and the nucleophilic cysteine by the other. The carboxy-terminal residues remain coordinated in the two active sites, predicting an inactive post-cleavage form. Proteolysis through formation of a composite active site occurs in the context of the viral polyprotein expressed in mammalian cells. These features offer unexpected insights into polyprotein processing by hepatitis C virus and new opportunities for antiviral drug design.

Crystallization of the catalytic domain of the hepatitis C virus (HCV) NS2-3 protease ($NS2^{pro}$, consisting of residues 94–217 of NS2; Fig. 1a) using native and selenomethionine-containing protein yielded two crystal forms with the same space group ($P2_1$, Supplementary Table 1). The asymmetric units of the native and selenomethionine-containing protein contained twelve and six $NS2^{pro}$ molecules organized into six and three tightly packed dimers, respectively (Supplementary Fig. 1).

The $NS2^{pro}$ monomer consists of two subdomains connected by an extended linker (Fig. 1b and Supplementary Fig. 2a). The amino-terminal subdomain contains two antiparallel α-helices (H1 and H2) followed by several turns and loops that contact both H1 and H2. The polypeptide chain continues into an extended region before entering a four-stranded, antiparallel β-sheet in the C-terminal subdomain. The last β-strand continues to the C terminus of NS2. Figure 1c, d represents views of the $NS2^{pro}$ dimer, which resembles a 'butterfly' with two-fold symmetry along the vertical axis (Fig. 1c). The N-terminal subdomain of one molecule interacts with the C-terminal subdomain of the other molecule and *vice versa*. The two extended linkers cross over in the middle of each molecule and each contribute a β-strand (b1) to the antiparallel β-sheet in the C-terminal subdomain of the other molecule. The N termini of the two monomers lie relatively close to each other, whereas the solvent-exposed C termini are positioned on opposite sides of the molecule.

Critical residues for NS2-3 proteolytic activity[7,8], His 143 and Glu 163, are located in the loop region following helix H2 in the N-terminal subdomain, whereas another critical residue, Cys 184, lies at the end of the linker arm in the b1–b2 loop of the C-terminal subdomain. At the dimer interface, the histidine and glutamate residues from one monomer are close to the cysteine from the other chain (Fig. 2a). The arrangement of these three residues is suggestive of a composite cysteine protease active site (Fig. 2b and Supplementary Fig. 2b).

The $NS2^{pro}$ structure represents a novel protein fold (DALI server[10] *Z* scores of less than 3.0). However, superimposing His 143, Glu 163 and Cys 184 of $NS2^{pro}$ with the active sites from cysteine proteases such as papain[11] (Fig. 2c) and poliovirus 3C protease[12] (Fig. 2d) demonstrated a similar spatial distribution. The orientation of the catalytic cysteine residues from $NS2^{pro}$ and $3C^{pro}$ is similar to the catalytic serine residues of Sindbis virus capsid[13] (Fig. 2e) and the cellular protease subtilisin[14] (Fig. 2f). Thus, like poliovirus $3C^{pro}$, HCV NS2-3 is a cysteine protease with a serine protease active site geometry[15]. To the best of our knowledge, $NS2^{pro}$ represents the first example of a cysteine or serine protease that forms a dimer containing a pair of composite active sites. However, these features are reminiscent of retroviral aspartic proteases, which consist of dimers with a single active site at the dimerization interface[16]. Other proteases such as caspases[17] require dimerization for activity, but they do not contain composite active sites.

Interestingly, Pro 164, which is entirely conserved in all HCV sequences, has a *cis*-peptide conformation (Fig. 2b and Supplementary Fig. 2c). Pro 164 may bend the peptide backbone of the catalytic Glu 163 to establish the correct geometry of the glutamate side chain for catalysis. In addition, the *cis*-proline may contribute to dimer stabilization as the linker connecting the two subdomains follows Pro 164.

The backbone carboxylic acid of the C-terminal residue of $NS2^{pro}$, Leu 217, remains coordinated in the active site via contacts to the side chains of His 143 and Cys 184, and to the backbone nitrogen of Cys 184 (Fig. 2b). Comparison of the $NS2^{pro}$ structure with protease–inhibitor complexes indicates that the C terminus of NS2 may exert an inhibitory function after cleavage[8]. Sindbis virus capsid protein mediates a single autoproteolytic cleavage at its C terminus, which remains bound to the active site inhibiting further catalysis[13]. The C-terminal residues of Sindbis virus capsid (Trp 264) and HCV NS2 (Leu 217) are in the same orientation relative to the catalytic triad (Fig. 2e). Comparing the structure of $NS2^{pro}$ with subtilisin bound to the inhibitor Eglin-C[14] demonstrates that the C terminus of $NS2^{pro}$ occupies a position equivalent to the inhibitor (Fig. 2f). We propose a model in which each $NS2^{pro}$ molecule catalyses a single NS2-3 cleavage event, allowing tightly regulated processing. However, we cannot rule out the possibility that NS2 mediates proteolysis of other

[1]Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.
*These authors contributed equally to this work.

viral or cellular proteins if the C-terminal β-strand (b5) is displaced from the active site.

Processing at the NS2/NS3 junction requires the NS3 protease domain and is stimulated by the addition of exogenous zinc[7,8,18,19]. However, because NS2 contains a complete cysteine-protease active site with the C terminus positioned for catalysis, the function of NS3 remains undetermined. The crystal structure of $NS2^{pro}$ represents the post-cleavage form, which may differ from the NS2-3 precursor. NS3 may interact with the highly conserved surface surrounding the NS2 active site (Supplementary Fig. 3), contributing to a functional catalytic environment and correct positioning of the scissile bond. The backbone nitrogen of Cys 184 contacts the carboxylic acid of Leu 217 and may serve as part of the oxyanion hole to stabilize the transition state during catalysis. A residue in uncleaved NS2-3 (either a backbone nitrogen of NS2 oriented differently in the pre-cleavage form or a residue within the NS3 serine protease domain) may also contribute to the oxyanion hole. The zinc requirement may be due to its structural function in NS3 (refs 5, 6) rather than a role in NS2-3 catalysis. Thus, limiting zinc could indirectly inhibit NS2-3 cleavage by affecting NS3 folding[20]. Consistent with this idea, the NS2-3 protease is inhibited by mutations in zinc-coordinating residues of NS3 (refs 7, 8).

Molecular surface analysis and biochemical data support the $NS2^{pro}$ dimer model. $NS2^{pro}$ shows a high degree of amino-acid sequence conservation at the interface between the two monomers (Supplementary Fig. 3). The cleavage rate of purified NS2-3 is concentration dependent, indicating that the active form of the protease is oligomeric[19]. Analytical ultracentrifugation of $NS2^{pro}$ yielded a single, monodisperse species with a molecular weight of 39 kDa that most likely corresponds to a dimer with bound detergent (data not shown). Moreover, cross-linking of $NS2^{pro}$ in solution with

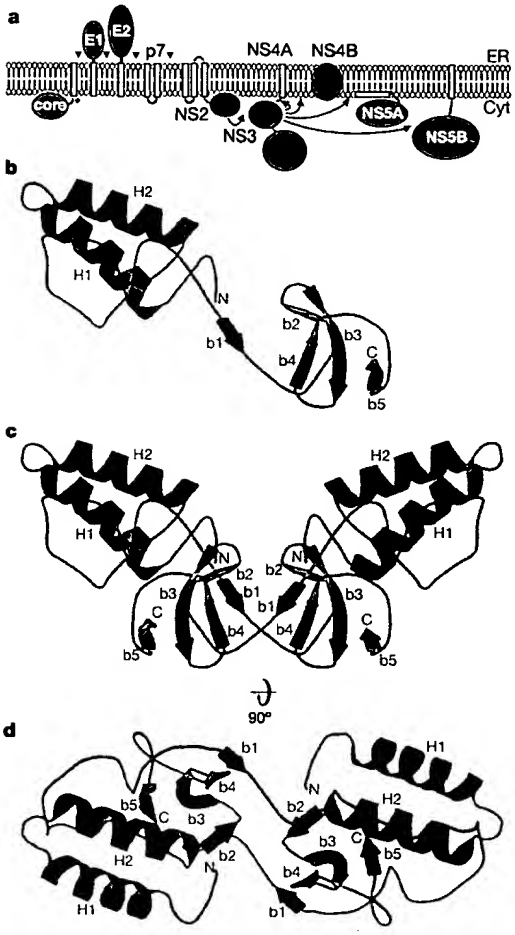


Figure 1 | Processing of the HCV polyprotein and architecture of $NS2^{pro}$. **a**, Membrane association of the HCV polyprotein, which contains a core, envelope proteins E1 and E2, p7, and nonstructural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A and NS5B. Cleavage in the structural region and at the N terminus of NS2 occurs by action of the host signal peptidase (filled arrowheads) and host signal peptide peptidase (diamond). The NS2-3 protease, which consists of residues 94–217 of NS2 (blue) and residues 1–181 of NS3 (green), cleaves at the NS2/NS3 junction (blue arrow). All cleavages downstream of NS3 are mediated by the NS3-4A protease (green arrows). The model shown here depicts NS2 with three N-terminal transmembrane segments followed by the protease domain on the cytoplasmic face of the endoplasmic reticulum (ER) membrane[19]. Alternatively, the NS2 C terminus may be localized to the ER lumen[28,29]. Cyt, cytosol. **b**, Ribbon diagram showing the NS2 monomer. The N and C termini, and secondary structure elements (helices H1 and H2; β-strands b1–b5), are labelled. **c**, Ribbon diagram showing the NS2 dimer with one monomer in blue, the other in red. The N and C termini, and secondary structure elements of each monomer, are indicated. **d**, Ribbon diagram of the NS2 dimer viewed perpendicular to helices H1 and H2. This view is a 90° rotation around the horizontal axis in **c**.

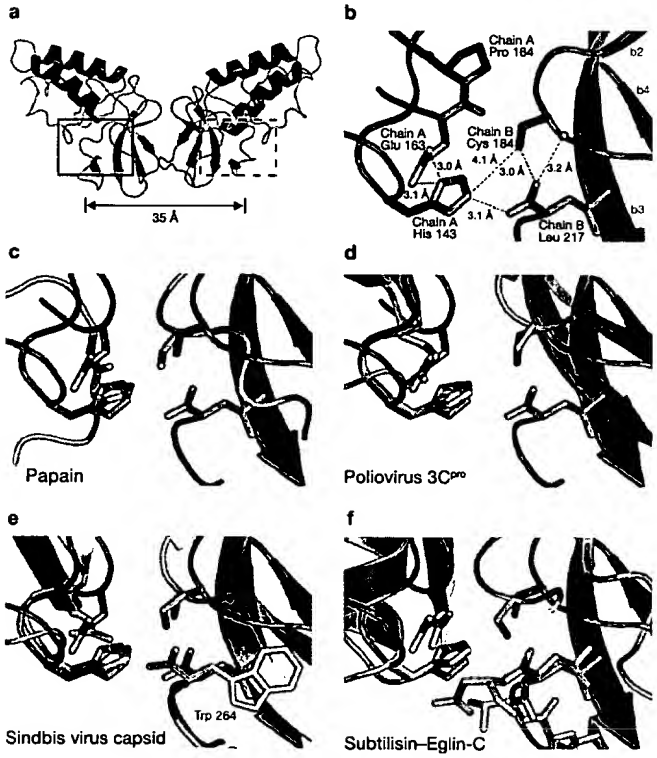


Figure 2 | The active site of NS2 and comparison with other proteases. **a**, Location of the two active sites in the NS2 dimer (boxed regions). The solid-lined box represents the active site displayed in **b**–**f**. The distance between the two active sites is indicated. **b**, The NS2 active site. Residues His 143, Glu 163, Pro 164, Cys 184 and Leu 217 are shown as stick drawings. The active site is composed of His 143 and Glu 163 from one molecule of the dimer (chain A, drawn in blue), and Cys 184 from the other molecule (chain B, drawn in red). The C-terminal residue, Leu 217, originates from the same chain as Cys 184. Dashed lines indicate contacts between selected residues. The length of each contact is provided. **c**–**f**, Superimposition of His 143, Glu 163 and Cys 184 of NS2 with similar catalytic sites of the proteases papain (**c**), poliovirus 3C protease (**d**), Sindbis virus capsid (**e**) and subtilisin bound to the inhibitor Eglin-C (coloured yellow) (**f**). The orientation and colouring of the NS2 active site is identical to **b**, with the other proteases shown in grey. In **e**, the C-terminal residue Trp 264 of Sindbis virus capsid is labelled.

disuccinimidyl suberate (DSS) led to the identification of a dimeric species (Supplementary Fig. 4).

A series of experiments in mammalian cells was designed to test whether $NS2^{pro}$ can form dimers with a functional composite active site *in vivo*. HCV full-length polyproteins containing either a H143A or a C184A mutation in the NS2 active site are defective in NS2-3 processing[7,8]. However, if a composite active site can form, co-expression of the two mutant polyproteins should result in partial NS2-3 cleavage (Fig. 3a). Indeed, when HCV polyproteins with NS2 containing either a H143A or C184A mutation were co-expressed, NS2 and NS3 cleavage products were detected (Fig. 3b), indicating the formation of a functional composite active site. Moreover, it is possible to predict from the crystal structure which mutant polypeptide in the mixing experiment is cleaved, because the C-terminal Leu 217 and the catalytic Cys 184 originate from the same chain. Thus, mixing of the two mutants should lead to cleavage of NS2-3(H143A), whereas NS2-3(C184A) is predicted to remain unprocessed. By expressing NS2-3 proteins with a Flag or haemagglutinin (HA) tag fused to the N terminus of NS2, it is possible to distinguish Flag–NS2 from HA–NS2 by immunoprecipitation using epitope-specific antibodies and by different electrophoretic mobilities of the various polypeptides (Fig. 3c). Mixing of Flag–NS2-3(H143A) with HA–NS2-3(C184A) resulted in cleavage of Flag–NS2-3, whereas HA–NS2-3 remained unprocessed (Fig. 3c, second lane from the right). When HA–NS2-3(H143A) was mixed with Flag–NS2-3(C184A), only HA–NS2-3 was cleaved (Fig. 3c, first lane from the right). Mixing wild-type Flag–NS2-3 with double-mutant HA–NS2-3(H143A/C184A) or *vice versa* yielded only cleaved wild-type NS2 (Fig. 3d), whereas the double-mutant polypeptide remained unprocessed because neither of the composite active sites is functional when a wild-type and a double-mutant NS2-3 dimerize. Finally, when cells are co-transfected with wild-type Flag–NS2-3 and HA–NS2-3 and lysed with a mild detergent, Flag–NS2 and HA–NS2 can be co-precipitated using either an anti-Flag or an anti-HA antibody (Fig. 3e). These data strongly support the $NS2^{pro}$ crystal structure and prove that NS2 can form dimers with composite functional active sites.

Our data change the current view of HCV polyprotein processing and raise interesting regulatory possibilities. Previously, NS2-3 cleavage was thought to occur as a unimolecular reaction in *cis*. The apparent requirement for dimerization to form an active NS2-3 protease suggests that NS2-3 cleavage and therefore formation of the active RNA replicase may be dependent on the concentration of NS2. Thus, a requirement for NS2-3 dimerization and subsequent proteolytic processing could delay the initiation of RNA replication, which may allow the virus to accumulate sufficient amounts of active NS3-4A protease to antagonize the induction of type I interferons by the host[21].

The cleaved NS2 dimer and higher-order oligomers may have additional roles in the viral life cycle, such as a function in membrane-associated virus assembly[22]. The solvent-accessible surface of the $NS2^{pro}$ dimer, coloured according to electrostatic potential (Fig. 4a, b), showed a high proportion of neutral and basic regions. The surface of the molecule near helices H1 and H2 is mainly hydrophobic, with basic residues lying underneath. The crystal structure of $NS2^{pro}$ contained several molecules of *n*-octyl-β-glucoside and *n*-decyl-β-maltoside interacting with those two helices (Fig. 4d, e and Supplementary Fig. 2d). We propose a model in which $NS2^{pro}$ interacts peripherally with cellular membranes (Fig. 4c, f). The N termini of two $NS2^{pro}$ monomers would lie

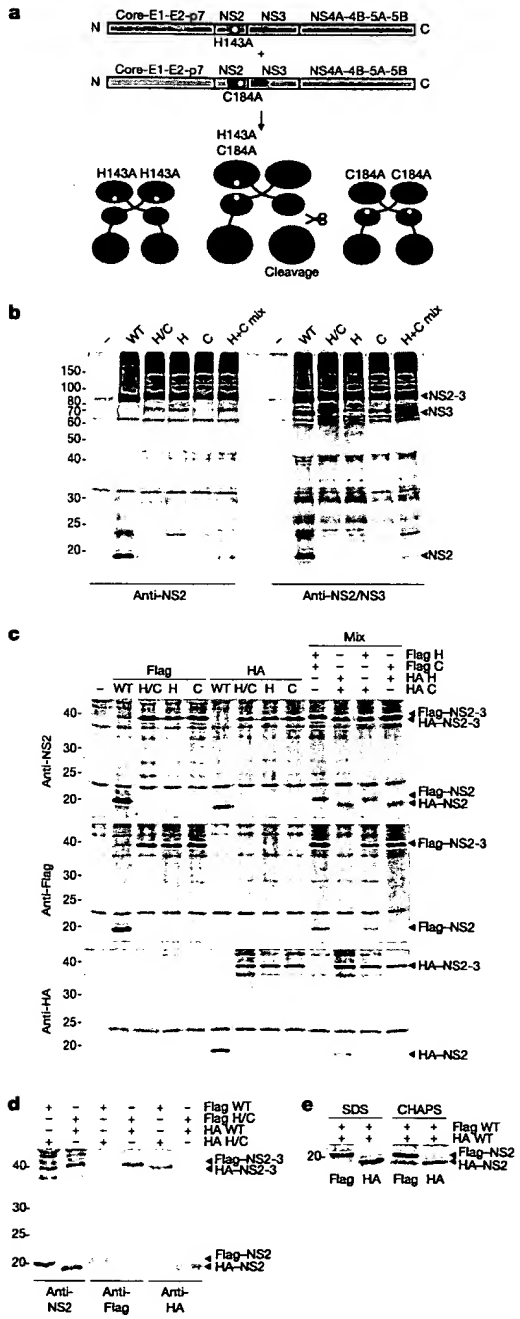


Figure 3 | Dimerization of NS2 and formation of a composite active site in mammalian cells. **a**, Mixing of two HCV polyproteins, each containing a point mutation in the NS2 active site (H143A or C184A; shown in blue and red, respectively) leads to the formation of NS2-3 mutant homo- and hetero-dimers. Dimerization between NS2(H143A) and NS2(C184A) yields one defective and one functional active site, resulting in partial NS2-3 cleavage. The protease domain of NS3 is shown in green. **b**, Mixing experiment in Huh-7.5 cells of HCV full-length polyproteins containing a wild-type (WT), single- (H or C) or double-mutant (H/C) NS2 active site, followed by metabolic labelling, cell lysis and immunoprecipitation using anti-NS2 or anti-NS2-3 antibodies. **c**, Co-transfection of U2OS cells with Flag- and HA-tagged NS2-3 constructs (consisting of residues 19–217 of NS2 and 1–181 of NS3) containing a wild-type or mutant NS2 active site as indicated on top of the lanes, followed by metabolic labelling, cell lysis and immunoprecipitation using anti-NS2, anti-Flag or anti-HA antibodies. **d**, Mixing experiment of Flag–NS2-3 and HA–NS2-3 containing a wild-type or double-mutant NS2 active site, as described above. **e**, Co-immunoprecipitation using anti-Flag or anti-HA antibodies of co-transfected Flag–NS2-3 and HA–NS2-3, lysed either in SDS or CHAPS buffer. Positions of the NS2, NS3 and NS2-3 proteins are shown by arrowheads on the right. Molecular weight markers are indicated on the left. H, H143A; C, C184A; H/C, H143A/C184A.

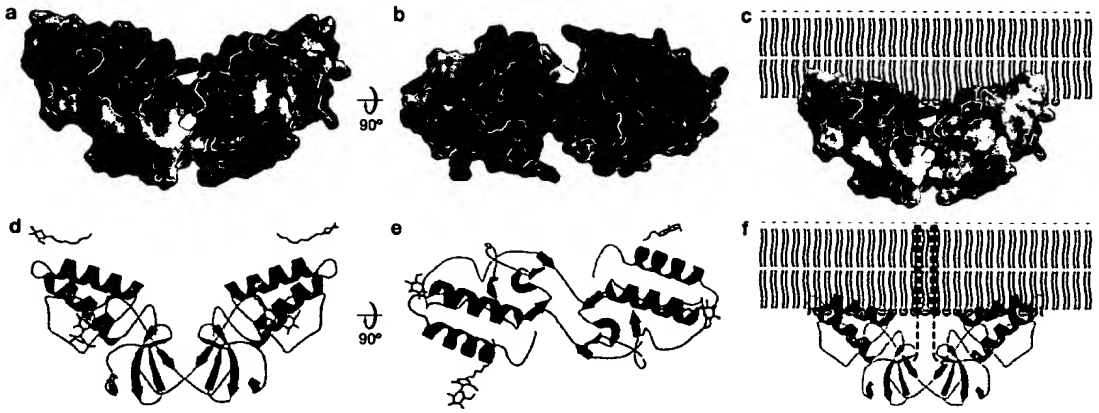


Figure 4 | Model for membrane association of NS2. **a**, Solvent-accessible surface of the NS2 dimer coloured according to electrostatic potential: acidic (red), neutral (white) and basic (blue). **b**, Electrostatic surface of the NS2 dimer rotated 90° around the horizontal axis shown in **a**. **c**, Electrostatic surface of the NS2 dimer inserted peripherally into a cellular membrane. **d**, Ribbon diagram of the NS2 dimer oriented as in **a**, with three *n*-octyl-β-glucoside molecules and one *n*-decyl-β-maltoside molecule bound to the N-terminal subdomains. **e**, Ribbon diagram of the NS2 dimer oriented as in **b**, showing the four bound detergent molecules. **f**, Ribbon diagram of the NS2 dimer positioned relative to the membrane. In this model, the hydrophobic part of helix H2 from each subunit is peripherally inserted into the lipid bilayer, with basic amino acids on the side of the helix involved in neutralizing the charge of polar lipid head groups in the membrane. The N-terminal transmembrane segments of NS2 (shown as dotted lines) would extend into the membrane.

close to the membrane (Fig. 4f). In the full-length protein, dimerization of NS2 may cause the N-terminal transmembrane domains of two monomers to form a 'bundle' of transmembrane segments that may serve an important function during virion morphogenesis.

The $NS2^{pro}$ structure presented here will allow further studies to elucidate the role of NS2 dimerization and other functions of NS2 in the viral life cycle. In addition, the structure establishes a foundation for the design of small-molecule inhibitors directed against the well-defined active site cleft and other conserved features of the protein.

METHODS

See Supplementary Information for detailed methods.

Protein preparation. $NS2^{pro}$ (NS2 residues 94–217), a truncated form of NS2 shown to be proteolytically active in the context of an NS2-3 precursor that includes the NS3 protease domain[18,19], was expressed in *Escherichia coli*. Lysis and subsequent purification of the protein by immobilized-metal affinity chromatography, ion exchange chromatography and gel filtration was done in the presence of detergent. The final concentration on a cation exchange column yielded highly pure protein at 6–9 mg ml^{-1}.

Crystal growth and freezing. Crystals of $NS2^{pro}$ were grown using hanging-drop vapour diffusion at 4 °C. The well contained a solution of 0.1 M Tris, pH 8.5, 0.8 M ammonium acetate, 0.25 M lithium chloride and 12% PEG 3350. Crystals were frozen in well solution supplemented with 5.4 mM decyl maltoside using stepwise addition of glycerol to a final concentration of 25%.

Data collection and structure determination. Data were collected at beamlines X9A and X29 at the Brookhaven National Laboratory's National Synchrotron Light Source. Phases were calculated from selenomethionine-containing protein by multiwavelength anomalous diffraction (MAD). After data indexing and scaling with DENZO/SCALEPACK[23], 19 of the 24 selenium sites were found using SnB[24]. An interpretable electron density map was obtained using MLPHARE[25], followed by density modification and phase combination using SOLOMON and DM[25]. Several rounds of iterative model building and refinement were done using O[26] and CNS[27]. The 2.9 Å resolution structure was used as a search model to obtain phases for a native data set at 2.28 Å resolution using molecular replacement. The final model contained 176 solvent molecules, 12 detergent molecules, and 12 molecules of $NS2^{pro}$. A summary of the refinement statistics is shown in Supplementary Table 1.

Expression in mammalian cells. NS2 containing wild-type or mutant active site residues was expressed in mammalian cells in the context of a full-length viral polyprotein or as NS2-3 precursor with an N-terminal Flag or HA tag. Cells metabolically labelled with ^{35}S were lysed and immunoprecipitation of NS2 performed using antibodies against NS2 or the Flag and HA tags. Proteins were separated by SDS–polyacrylamide gel electrophoresis and visualized by autoradiography.

Received 8 May; accepted 9 June 2006.
Published online 23 July 2006.

1. Alter, H. J. & Seeff, L. B. Recovery, persistence, and sequelae in hepatitis C virus infection: a perspective on long-term outcome. *Semin. Liver Dis.* 20, 17–35 (2000).
2. Hoofnagle, J. H. Course and outcome of hepatitis C. *Hepatology* 36, S21–S29 (2002).
3. Bartenschlager, R., Frese, M. & Pietschmann, T. Novel insights into hepatitis C virus replication and persistence. *Adv. Virus Res.* 63, 71–180 (2004).
4. Lindenbach, B. D. & Rice, C. M. in *Fields Virology* (eds Knipe, D. M. *et al.*) 991–1041 (Lippincott Williams & Wilkins, Philadelphia, Pennsylvania, 2001).
5. Love, R. A. *et al.* The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* 87, 331–342 (1996).
6. Kim, J. L. *et al.* Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87, 343–355 (1996).
7. Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. A second hepatitis C virus-encoded proteinase. *Proc. Natl Acad. Sci. USA* 90, 10583–10587 (1993).
8. Hijikata, M. *et al.* Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J. Virol.* 67, 4665–4675 (1993).
9. Wu, Z., Yao, N., Le, H. V. & Weber, P. C. Mechanism of autoproteolysis at the NS2–NS3 junction of the hepatitis C virus polyprotein. *Trends Biochem. Sci.* 23, 92–94 (1998).
10. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138 (1993).
11. Pickersgill, R. W., Harris, G. W. & Garman, E. Structure of monoclinic papain at 1.60 Å resolution. *Acta Crystallogr. B* 48, 49–66 (1992).
12. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* 273, 1032–1047 (1997).
13. Tong, L., Wengler, G. & Rossmann, M. G. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *J. Mol. Biol.* 230, 228–247 (1993).
14. McPhalen, C. A. & James, M. N. Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* 27, 6582–6598 (1988).
15. Malcolm, B. A. The picornaviral 3C proteinases: cysteine nucleophiles in serine proteinase folds. *Protein Sci.* 4, 1439–1445 (1995).
16. Wlodawer, A. & Gustchina, A. Structural and biochemical studies of retroviral proteases. *Biochim. Biophys. Acta* 1477, 16–34 (2000).
17. Boatright, K. M. & Salvesen, G. S. Mechanisms of caspase activation. *Curr. Opin. Cell Biol.* 15, 725–731 (2003).
18. Thibeault, D., Maurice, R., Pilote, L., Lamarre, D. & Pause, A. *In vitro* characterization of a purified NS2/3 protease variant of hepatitis C virus. *J. Biol. Chem.* 276, 46678–46684 (2001).
19. Pallaoro, M. *et al.* Characterization of the hepatitis C virus NS2/3 processing reaction by using a purified precursor protein. *J. Virol.* 75, 9939–9946 (2001).

20. Urbani, A. *et al.* The metal binding site of the hepatitis C virus NS3 protease. A spectroscopic investigation. *J. Biol. Chem.* **273**, 18760–18769 (1998).
21. Gale, M. Jr & Foy, E. M. Evasion of intracellular host defence by hepatitis C virus. *Nature* **436**, 939–945 (2005).
22. Pietschmann, T. *et al.* Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc. Natl Acad. Sci. USA* **103**, 7408–7413 (2006).
23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
24. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Crystallogr.* **32**, 120–124 (1999).
25. Collaborative Computational Project No. 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Santolini, E., Pacini, L., Fipaldini, C., Migliaccio, G. & Monica, N. The NS2 protein of hepatitis C virus is a transmembrane polypeptide. *J. Virol.* **69**, 7461–7471 (1995).
29. Yamaga, A. K. & Ou, J. H. Membrane topology of the hepatitis C virus NS2 protein. *J. Biol. Chem.* **277**, 33228–33234 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank R. MacKinnon, S. Darst and H. Mueller for the use of X-ray diffractometers, related equipment and software. We appreciate access to beamlines X9A and X29 of the National Synchrotron Light Source (NSLS) at the Brookhaven National Laboratory and acknowledge the assistance of the NSLS staff. We wish to thank T. Tellinghuisen for assistance with data collection. D. Jeruzalmi provided the program msf_similarity_to_pdb and LOOSENGRASP. We would like to thank S. Burley, S. Darst, L. Dustin, M. Evans, C. Jones, C. Murray, B. Lindenbach, G. Randall and T. Tellinghuisen for input on the manuscript, and K. Kuo and S. You for help with artwork. I.C.L. was supported by fellowships from the Swiss National Science Foundation, the Roche Research Foundation and the Swiss Society for Medical-Biological Grants. J.M. was supported as a Merck Fellow of the Life Sciences Research Foundation. Additional financial support for this work came from grants from the National Institutes of Health and the Greenberg Medical Research Institute (C.M.R.).

Author Contributions I.C.L., J.M. and C.M.R. conceived the experiments. I.C.L. generated all reagents, materials, proteins and crystals with assistance from J.M. All data collection and processing was done by J.M. and I.C.L. Model building and refinement were performed by I.C.L. and J.M. Cross-linking and mammalian expression studies were carried out by I.C.L. and T.G.D. The manuscript was written by I.C.L., J.M. and C.M.R.

Author Information The atomic coordinates for this structure have been deposited in the Protein Data Bank under accession code 2HD0. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.M.R. (ricec@rockefeller.edu) or J.M. (marcotj@rockefeller.edu).

JOURNAL OF VIROLOGY, Dec. 1994, p. 8147–8157
0022-538X/94/$04.00+0

Vol. 68, No. 12

Hepatitis C Virus NS3 Serine Proteinase: *trans*-Cleavage Requirements and Processing Kinetics

CHAO LIN, BÉLA M. PRÁGAI, ARASH GRAKOUI, JIAN XU, AND CHARLES M. RICE*

Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110-1093

Received 1 August 1994/Accepted 16 September 1994

The hepatitis C virus H strain (HCV-H) polyprotein is cleaved to produce at least 10 distinct products, in the order of NH_2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH. An HCV-encoded serine proteinase activity in NS3 is required for cleavage at four sites in the nonstructural region (3/4A, 4A/4B, 4B/5A, and 5A/5B). In this report, the HCV-H serine proteinase domain (the N-terminal 181 residues of NS3) was tested for its ability to mediate *trans*-processing at these four sites. By using an NS3-5B substrate with an inactivated serine proteinase domain, *trans*-cleavage was observed at all sites except for the 3/4A site. Deletion of the inactive proteinase domain led to efficient *trans*-processing at the 3/4A site. Smaller NS4A-4B and NS5A-5B substrates were processed efficiently in *trans*; however, cleavage of an NS4B-5A substrate occurred only when the serine proteinase domain was coexpressed with NS4A. Only the N-terminal 35 amino acids of NS4A were required for this activity. Thus, while NS4A appears to be absolutely required for *trans*-cleavage at the 4B/5A site, it is not an essential cofactor for serine proteinase activity. To begin to examine the conservation (or divergence) of serine proteinase-substrate interactions during HCV evolution, we demonstrated that similar *trans*-processing occurred when the proteinase domains and substrates were derived from two different HCV subtypes. These results are encouraging for the development of broadly effective HCV serine proteinase inhibitors as antiviral agents. Finally, the kinetics of processing in the nonstructural region was examined by pulse-chase analysis. NS3-containing precursors were absent, indicating that the 2/3 and 3/4A cleavages occur rapidly. In contrast, processing of the NS4A-5B region appeared to involve multiple pathways, and significant quantities of various polyprotein intermediates were observed. NS5B, the putative RNA polymerase, was found to be significantly less stable than the other mature cleavage products. This instability appeared to be an inherent property of NS5B and did not depend on expression of other viral polypeptides, including the HCV-encoded proteinases.

Hepatitis C viruses (HCVs) have recently been recognized as agents of the parentally transmitted form of non-A, non-B hepatitis (17, 41). Virtual elimination of HCV-contaminated blood has greatly reduced the incidence of posttransfusion hepatitis; however, HCV remains responsible for a significant proportion of community-acquired hepatitis (1). In most cases, HCV is not cleared and establishes a chronic infection that can be associated with chronic hepatitis and more severe liver disease such as cirrhosis and hepatocellular carcinoma (63). For these reasons, there is considerable interest in developing additional HCV-specific antiviral agents that can complement currently available alpha interferon therapy, which effectively controls disease in only a minority of HCV-infected patients.

At least 15 full-length HCV genome sequences, as well as partial sequences for many other isolates, have been reported (see reference 60 and citations therein). These data indicate the existence of multiple genotypes that can diverge by as much as 50% at the amino acid level (10, 64, 65). This group of related viruses is now classified as a separate genus in the family *Flaviviridae* (27), which includes two other genera, *Flavivirus* (12) and *Pestivirus* (20). The positive-strand HCV genome RNA is approximately 9.4 kb in length and contains a highly conserved 5′ noncoding region followed by a long open reading frame encoding a polyprotein of 3,010 to 3,033 amino acids (36, 51). Because a cell culture system supporting efficient HCV replication is lacking, efforts to define potential HCV-encoded polypeptides have utilized expression of HCV cDNA in cell-free translation and in cell cultures. The HCV polyprotein appears to be cleaved at multiple sites to produce at least 10 structural and nonstructural (NS) proteins (47). The order and nomenclature of these cleavage products for the HCV H strain (HCV-H) are NH_2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH, where C, E1, and E2 are putative structural proteins and the remaining NS proteins are believed to be replicase components (30–32, 47). Host signal peptidase in the endoplasmic reticulum lumen appears to catalyze cleavages in the structural-NS2 region (C/E1, E1/E2, E2/p7, and p7/NS2 sites) (33, 47), whereas an HCV-encoded serine proteinase located in the N-terminal one-third of the NS3 protein is responsible for four cleavages in the NS region (3/4A, 4A/4B, 4B/5A, and 5A/5B sites) (5, 22, 30, 34, 50, 69). Autocatalytic cleavage at the 2/3 site is mediated by a second HCV-encoded proteinase that encompasses the NS2 region and the NS3 serine proteinase domain (31, 35).

In this study, we tested the ability of the NS3 serine proteinase domain (called $NS3_{181}$) to mediate *trans*-processing at each of the four downstream sites. All four sites could be cleaved in *trans*; however, requirements for *trans*-cleavage varied for different sites. *trans*-cleavage at the 3/4A site was very inefficient, if there was any, when the substrate contained an inactivated serine proteinase. Coexpression of NS4A is required for cleavage at the 4B/5A site, but not at the 5A/5B site. We also tested the ability of the serine proteinases from two HCV subtypes (H and BK strains) to mediate *trans*-processing of heterologous HCV polyprotein substrates. Finally, we used a vaccinia virus recombinant expressing the entire HCV polyprotein to examine the processing kinetics in

* Corresponding author. Mailing address: Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110-1093. Fax: (314) 362-1232. Electronic mail address: rice@borcim.wustl.edu.

the NS region and the stability of HCV precursors and cleavage products.

MATERIALS AND METHODS

Cell cultures. The BHK-21 and CV-1 cell lines were obtained from the American Type Culture Collection, and the BSC-40 cell line (9) was obtained from D. Hruby (Oregon State University). Cell monolayers were grown in Eagle's minimal essential medium (MEM) supplemented with 2 mM L-glutamine, nonessential amino acids, penicillin, streptomycin, and 10% fetal bovine serum (FBS). The A16 subclone of the human hepatoma HepG2 cell line, generously provided by Alan Schwartz (Washington University), was maintained in Dulbecco's modified Eagle medium supplemented with penicillin, streptomycin, and 10% FBS.

Plasmid constructions. Standard recombinant DNA techniques (61) were used for construction of the expression plasmids described below. For all plasmids, regions of HCV-H coding sequence amplified by PCR were verified by DNA sequence analysis.

Synthetic oligonucleotides and PCR were used to engineer initiation or termination codons as well as convenient restriction sites for subcloning (5′ *Nco*I and 3′ *Xho*I sites) for several HCV-H expression constructs. These constructs (with the encoded polyproteins given in parentheses) are as follows (Fig. 1): pTM3/HCV1027-1657 (NS3), pBRTM/HCV1027-1711 (NS3-4A), pTM3/HCV1658-1711 (NS4A), pTM3/HCV1658-1972 (NS4A-4B), pTM3/HCV1658-2420 (NS4A-5A), pTM3/HCV1712-2420 (NS4B-5A), pBRTM/HCV1712-3011 (NS4B-5B), pBRTM/HCV1973-3011 (NS5A-5B), and pTM3/HCV 2421-3011 (Met-NS5B). The sequences encompassing the engineered initiation codons (boldface) are as follows (HCV-H sequence underlined): NS3, 5′-CCATGGCGCCC-3′; NS4A, 5′-CCATGGCCAGCACC-3′; NS4B, 5′-CCATGGCGTCTC AG-3′; NS5A, 5′-CCATGGGATCCGGC-3′; and NS5B, 5′-CCATGGGCTCAATG-3′. For the engineered termination codons (boldface), the surrounding sequences are as follows (HCV-H sequence underlined): NS3, 5′-GTCACGTGACTC GAG-3′; NS4A, 5′-GAGTGCTAGCTCGAG-3′; NS4B, 5′-CCATGCTAGCTCGAG-3′; and NS5A, 5′-TGCTGCTAGC TCGAG-3′.

pTM3/Ubiquitin-HCV2421-3011 (Ubi-NS5B) was constructed by ligation of two PCR-derived fragments into pTM3/HCV2421-3011 (Met-NS5B). The initiating methionine of the ubiquitin monomer corresponds to the ATG in the *Nco*I site of pTM3. The ubiquitin (double underlined)-NS5B (underlined) junction was created by using a *Bam*HI restriction site (boldface) as follows: CGC GGT GGA TCC ATG TCT. The template for PCR amplification of the ubiquitin cassette was pTM3/Ub-nsP4 (Tyr) (44).

Additional HCV-H expression plasmids (with the encoded polyproteins given in parentheses) were constructed by subcloning appropriate fragments from previously described constructs (Fig. 1). pTM3/HCV1027-1207 ($NS3_{181}$) was derived from pTM3/HCV1027-1657 (described above) and pTM3/HCV827-1207 (31); pTM3/HCV1027-1676 ($NS3\text{-}4A_{19}$) was derived from pTM3/HCV1027-1657 and pBRTM/HCV1-1676 (32); pTM3/HCV1027-1692 ($NS3\text{-}4A_{35}$) was derived from pTM3/HCV1027-1657 and pBRTM/HCV1-1692 (32); pBRTM/HCV1027-3011 $S_{1165}A$ (NS3-5B*) was derived from pBRTM/HCV1-3011 $S_{1165}A$ (30) and pBRTM/HCV1027-1711; pTM3/HCV1193-1657 ($NS3_{167\text{-}631}$) was derived from pBRTM/HCV1193-3011 (30) and pTM3/HCV1027-1657; and pTM3/HCV1193-1711 ($NS3_{167}\text{-}4A$) was derived from pBRTM/HCV 1193-3011, pBRTM/HCV1027-1711, and pTM3. pTM3/HCV

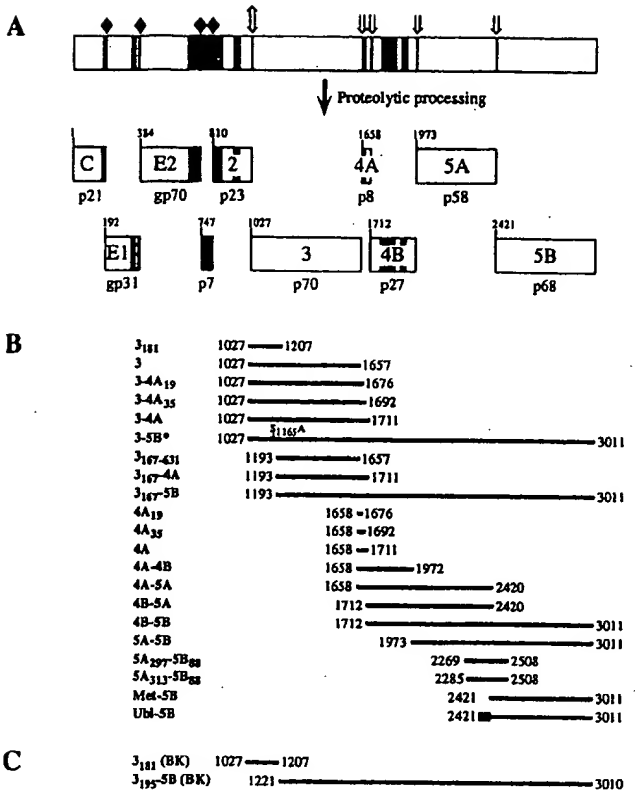


FIG. 1. HCV genome structure and expression constructs. (A) Diagram of the HCV-H strain polyprotein and its cleavage products shown as boxes. The identities of the mature proteins, including C, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B, are indicated (32, 47). The number at the top of each cleavage product indicates the position of its N-terminal residue in the polyprotein sequence. The apparent molecular masses for HCV proteins (p) and glycoproteins (gp) are indicated under each product (in kilodaltons). Regions containing predominantly uncharged amino acids are indicated as black bars. Also shown are putative cleavage sites for host signal peptidase (◆) (33, 47), the HCV NS2-3 proteinase (⇳) (31, 34), and the NS3 serine proteinase (⇓) (5, 22, 30, 34, 50, 69). (B) HCV-H polypeptide expression constructs used in this study. HCV polypeptide sequences present in each pBRTM/HCV or pTM3/HCV construct are indicated by black lines, which are drawn to scale and oriented with respect to the diagram of the HCV-H polyprotein. Numbers at the ends of each line refer to the first and last amino acids of the HCV polypeptide expressed by the particular construct. For simplicity, the NS prefix is not used for the nomenclature of each encoded polypeptide, which is indicated on the left. (C) HCV-BK polypeptide expression constructs. (See the legend to panel B for details.)

1658-1676 ($NS4A_{19}$) was constructed by deleting the *Hin*cII-*Nhe*I fragment of pTM3/HCV1658-1711 (the *Nhe*I site was filled in by using T4 DNA polymerase prior to ligation). pTM3/HCV1658-1692 ($NS4A_{35}$) was generated by deleting the *Nae*I-*Nhe*I fragment of pTM3/HCV1658-1711 (the *Nhe*I site was filled in by using T4 DNA polymerase prior to ligation). pTM3/HCV2269-2508 ($NS5A_{297}\text{-}5B_{88}$) was made by subcloning the 1,274-bp *Bsa*I-*Bgl*II fragment from pTM3/HCV1-2508 (32) into pTM3 digested with *Nco*I and *Bgl*II (the *Bsa*I and *Nco*I sites were filled in by T4 DNA polymerase prior to ligation). pTM3/HCV2285-2508 ($NS5A_{313}\text{-}5B_{88}$) was constructed by subcloning the 1,227-bp *Apa*I-*Bgl*II fragment of pTM3/HCV1-2508 (32) into pTM3, which had been previously

digested with *Nco*I and *Bgl*II (the *Apa*I and *Nco*I cleavage sites were trimmed and filled in, respectively, by T4 DNA polymerase prior to ligation).

Expression constructs for the HCV BK strain (HCV-BK) were made with cDNA clones generously provided by H. Okayama and A. Takamizawa (67). pTM3/HCV-BK1027-1207 [encoding polypeptide $NS3_{181}$(BK)] was constructed by subcloning a PCR fragment amplified from pUC19/BK-146 (67) into pTM3. The sequences encompassing the engineered initiation and termination codons (boldface) include (HCV-BK sequences underlined) an *Nco*I site at the N terminus (5'-CCATGGCTCCC-3') and a *Bam*HI site at the C terminus (5'-CGGTCTTAATAGGATCC-3'). pBRTM/HCV-BK1221-3011 was produced by subcloning appropriate fragments from four HCV-BK cDNA clones, including pUC19/BK-102, BK-112-1, BK-112-5, and BK-166. Because the HCV-BK coding sequence in pUC19/BK-102 clones was fused in frame to the AUG codon in the *Nco*I site of the adaptor sequence, pBRTM/HCV-BK1221-3011 encodes a polyprotein [$NS3_{195}$-5B(BK)] encompassing HCV-BK residues 1221 to 3011 after the initiating methionine.

Generation and growth of vaccinia virus-HCV recombinants. vHCV1027-1207 was generated by marker rescue of pTM3/HCV1027-1207 (49). Recombinant viruses were plaque purified three times under *gpt* selection (25) prior to growth of large-scale stocks. A vaccinia virus-HCV recombinant encoding the entire HCV-H open reading frame, vHCV1-3011, has been described previously (47). Stocks of vHCV1027-1207, vHCV1-3011, and vTF7-3, a vaccinia virus recombinant expressing the T7 DNA-dependent RNA polymerase (28), were grown in BSC-40 monolayers and partially purified (37), and titers of infectious progeny were determined by plaque assay on BSC-40 cells (37).

Transient expression with the vaccinia virus-T7 hybrid system. For expression assays utilizing vaccinia virus-HCV recombinants, monolayers of HepG2-A16 or BHK-21 cells in 35-mm-diameter dishes were infected with vTF7-3 alone or in combination with vHCV1-3011, vHCV827-3011 (32), or vHCV1027-1207. The multiplicity of infection for each recombinant was 10 PFU per cell. After adsorption for 60 min at room temperature, the inoculum was removed and replaced with MEM containing 2% FBS. Expression assays of transfected plasmid constructs utilized subconfluent monolayers of BHK-21 cells that had been previously infected with vTF7-3 as described above. Some of them were also coinfected with vHCV1027-1207. After removal of the inoculum, cells were transfected for 2 h at 37°C with a mixture consisting of 1 μg of plasmid DNA and 10 μg of Lipofectin (Bethesda Research Laboratories) in 0.5 ml of MEM. If two constructs were used in a single transfection, the amount of each plasmid varied from 0.5 μg to 1 μg, with a total of 1.5 μg of DNA mixed with 15 μg of Lipofectin.

For pulse-chase experiments, monolayers were washed once with prewarmed methionine-deficient MEM at 3 h postinfection and incubated in the same medium for 20 min at 37°C. Cells were labeled by incubation for 20 min at 37°C with methionine-deficient MEM supplemented with 100 μCi of ^{35}S-protein labeling mixture (NEN) per ml. For chase experiments, the labeling mixture was replaced with MEM containing 2% FBS, 1.5 mg of methionine per ml, and 100 μg of cycloheximide per ml and incubated for the indicated periods at 37°C. For steady-state labeling, cell monolayers were washed once at 3 h postinfection as described above and then were incubated for 4 h at 37°C with MEM containing 1/40th the normal concentration of methionine and cysteine, 2% FBS, and 40 μCi of ^{35}S-protein labeling mixture per ml.

Cell lysis, immunoprecipitation, and protein analyses. After labeling, cell monolayers were washed with phosphate-buffered saline and lysed with a solution of 0.5% sodium dodecyl sulfate (SDS), 50 mM Tris-Cl (pH 7.4), 1 mM EDTA, and 20 μg of phenylmethylsulfonyl fluoride per ml (0.3 ml/10^6 cells). Cellular DNA was sheared by repeated passage through a 27.5-gauge needle. Prior to immunoprecipitation (47), lysates were heated to 70°C for 10 min. Portions of each lysate were incubated either with 5 to 10 μl of the indicated rabbit polyclonal antisera or with 2 μl of serum JHF from an HCV-positive patient (32). Immune complexes were collected by using *Staphylococcus aureus* Cowan I (Calbiochem) as described previously (59), solubilized, and analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) (42) or Tricine-SDS-PAGE (62). After treatment for fluorography with En^{3}Hance (DuPont), gels were dried and exposed at −70°C with prefogged (43) X-ray film (Kodak). ^{14}C-methylated molecular weight marker proteins were purchased from Amersham.

Cell-free translation. The 5'-uncapped RNA transcripts were synthesized from linearized cDNA templates with T7 DNA-dependent RNA polymerase (Epicenter) (58). Cell-free translation mixtures with rabbit reticulocyte lysates (Promega) and [^{35}S]methionine (Amersham), were incubated for 1 h at 30°C essentially according to the manufacturer's instructions. The translation reactions were terminated by the addition of RNase A (Boehringer Mannheim) to 10 μg/ml, cycloheximide to 0.3 mg/ml, and cold methionine to 1 mM. A portion of the translation reaction mixtures was removed at the indicated time, diluted 10-fold with the Laemmli sample buffer, heated for 5 min at 95°C, and analyzed by SDS-PAGE as described above.

RESULTS

***trans*-Cleavage at all four serine proteinase-dependent sites.** The serine proteinase domain of HCVs was initially identified on the basis of sequence homology to members of the trypsin superfamily (7, 29). The predicted domain is approximately 180 residues and corresponds to the N-terminal one-third of NS3. This enzyme is required for processing in the NS3-4-5 region of the HCV polyprotein, and alanine substitutions for predicted active site residues (His-1083 or Ser-1165 for HCV-H) abolish cleavage at the 3/4A, 4A/4B, 4B/5A, and 5A/5B sites (5, 22, 30, 34, 50, 69). To purify and characterize this enzyme, we have used the vaccinia virus-T7 hybrid expression system to examine the ability of the predicted serine proteinase domain, expressed as an individual polypeptide ($NS3_{181}$), to mediate *trans*-cleavage at each of these four sites.

The first substrate examined was an NS3-5B polyprotein containing the Ala substitution at Ser-1165 (NS3-5B*) (Fig. 1). This mutation completely inactivates the serine proteinase, and no processed products were observed (Fig. 2). When coexpressed with $NS3_{181}$, cleavage occurred at the 4A/4B, 4B/5A, and 5A/5B sites, as evidenced by the appearance of NS4B (Fig. 2B), NS5A (Fig. 2C), and NS5B (Fig. 2D). In contrast, we observed a more-slowly-migrating NS3-specific product, presumably NS3-4A, in addition to a very faint band corresponding to NS3 (Fig. 2A). This suggests that very inefficient, if any, *trans*-cleavage occurred at the 3/4A site of this substrate.

The lack of *trans*-cleavage at the 3/4A site has been observed in other studies and has led to the proposal that this site can only be cleaved in *cis* (5, 69). However, all substrates examined thus far contained an inactivated NS3 serine proteinase domain, which might interfere with the accessibility of the 3/4A

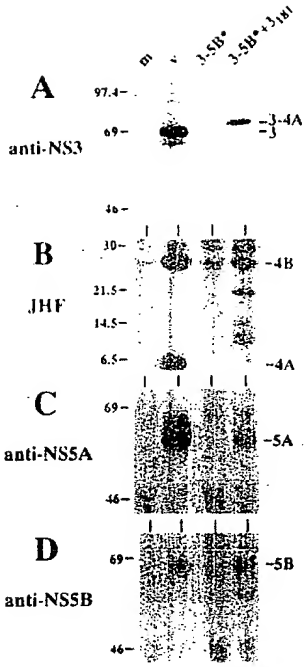


FIG. 2. *trans*-Processing of the HCV-H NS3-5B* polyprotein. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). Some monolayers were also transfected with pBRTM/HCV1027-3011 $S_{1165}A$ (3-5B*). Cells were metabolically labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. Cell lysates were immunoprecipitated with the following HCV-specific antisera: NS3-specific WU117 (A), NS5A-specific WU123 (C), NS5B-specific WU115 (D), or human patient serum JHF (B). It should be noted that the NS3 serine proteinase domain is not recognized by either human patient serum JHF or rabbit antiserum WU117, which was raised against the NS3 helicase domain. Immunoprecipitated proteins were solubilized and separated by electrophoresis on 8% (A, C, and D) or 14% (B) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

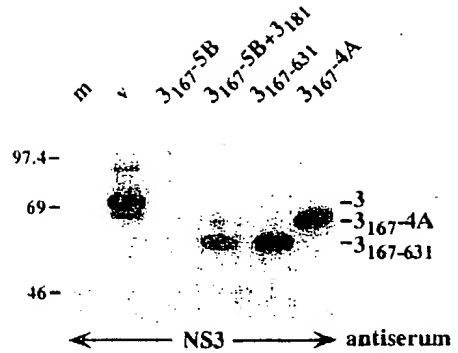


FIG. 3. Requirements for *trans*-cleavage at the 3/4A site. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). Some monolayers were also transfected with pBRTM/HCV1193-3011 (3_{167}-5B), pTM3/HCV1193-1657 ($3_{167-631}$), or pTM3/HCV1193-1711 (3_{167}-4A). Cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV NS3-specific products were immunoprecipitated with rabbit antiserum WU117, solubilized, and analyzed by SDS-PAGE (8% polyacrylamide). HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

site for *trans*-cleavage. To test this possibility, we expressed a polyprotein, $NS3_{167}$-5B, which begins with residue 167 of NS3 and therefore lacks the majority of the serine proteinase domain. Marker proteins were also expressed beginning with NS3 residue 167 and extending to the C terminus of NS3 ($NS3_{167-631}$) or NS4A ($NS3_{167}$-4A) (Fig. 1). Processed products were not observed when $NS3_{167}$-5B was expressed alone (Fig. 3). During coexpression with $NS3_{181}$, two NS3-specific cleavage products were observed: a major product comigrating with $NS3_{167-631}$ and traces of a larger species comigrating with $NS3_{167}$-4A (Fig. 3). These results clearly demonstrate that $NS3_{181}$ can mediate efficient *trans*-cleavage at the 3/4A site of a substrate which lacks the inactivated proteinase domain.

In contrast to the flaviviruses, where NS2B is absolutely required for NS3 serine proteinase activity (11, 24, 57), HCV sequences upstream of NS3 are not required for serine proteinase-dependent cleavages (5, 22, 30). However, the potential role of downstream viral polypeptide sequences in proteolysis has not been examined. To address this possibility, we tested *trans*-cleavage of NS4A-4B, NS4B-5A, and NS5A-5B substrates, each of which contained only a single proteinase-dependent cleavage site (Fig. 1). When expressed alone, only the appropriate unprocessed polyproteins were present (Fig. 4). When coexpressed with $NS3_{181}$, NS4A-4B was processed to yield NS4A and NS4B (Fig. 4A), and NS5A-5B yielded NS5A and NS5B (Fig. 4C). To develop shorter substrates convenient for in vitro proteinase assays, we examined *trans*-processing of $NS5A_{297}$-$5B_{88}$ and $NS5A_{313}$-$5B_{88}$, which contain the C-terminal 152 and 136 residues of NS5A, respectively, followed by the N-terminal 88 amino acids of NS5B (Fig. 1). $NS5_{297}$-$5B_{88}$ was processed efficiently by $NS3_{181}$ as evidenced by the conversion of most of $NS5A_{297}$-$5B_{88}$ to $NS5A_{297-448}$. Nearly complete *trans*-cleavage at the 5A/5B site was also observed for $NS5A_{313}$-$5B_{88}$ (Fig. 4D). These results indicate that only limited flanking sequences are necessary for efficient *trans*-cleavage at the 5A/5B site by the $NS3_{181}$ serine proteinase. Since these substrates do not overlap, these data exclude an absolute requirement for one of the downstream viral polypeptides for serine proteinase activity. In contrast to the results with the NS4A-4B, NS5A-5B, and NS3-5B* substrates, however, no *trans*-cleavage of NS4B-5A was observed (Fig. 4B).

NS4A is required for cleavage at the 4B/5A site. Since *trans*-cleavage at the 4B/5A site occurred for the NS3-5B* substrate but not for NS4B-5A, we examined *trans*-cleavage of NS4A-5A and NS4B-5B polyprotein substrates (Fig. 1). When coexpressed with $NS3_{181}$, the 4B/5A cleavage occurred in the NS4A-5A substrate (Fig. 5A, lane 4) but not in NS4B-5B (data not shown). These results suggested that, in addition to the $NS3_{181}$ proteinase domain, NS4A was required for cleavage at the 4B/5A site. The requirement for NS4A was strengthened by the observation that processing of NS4B-5A was restored by coexpression of NS4A in *trans*. Cleavage at the 4B/5A site of this substrate occurred when NS4A was expressed either as part of the proteinase (NS3-4A) (Fig. 5A, lane 9) or as an individual polypeptide (NS4A) together with $NS3_{181}$ (Fig. 5A, lane 7). These results clearly demonstrate that NS3-mediated cleavage at the 4B/5A site requires NS4A, a small protein of 54 amino acids with a hydrophobic N-terminal half and a C-terminal half rich in charged residues (see Discussion).

Since NS3-4A was fully active for *trans*-cleavage at the 4B/5A site, we made two constructs with C-terminal deletions

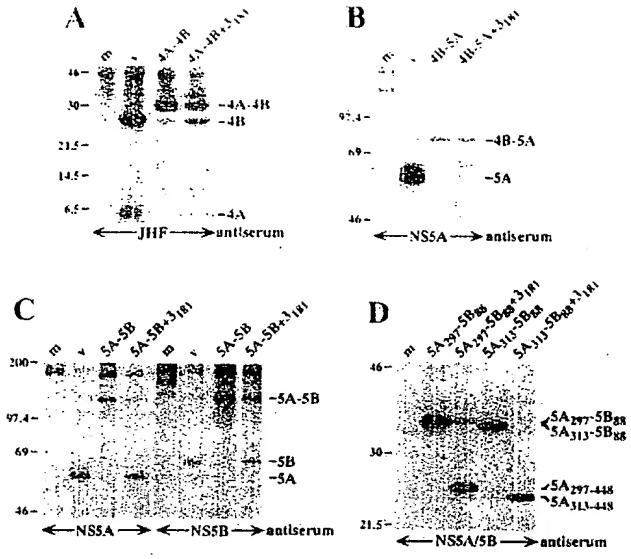


FIG. 4. *trans*-Processing of HCV-H polyproteins containing only one serine proteinase-dependent site. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). As indicated, some monolayers were also transfected with pTM3/HCV1658-1972 (4A-4B), pTM3/HCV1712-2420 (4B-5A), pBRTM/HCV1973-3011 (5A-5B), pTM3/HCV2269-2508 ($5A_{297}$-$5B_{88}$), or pTM3/HCV2285-2508 ($5A_{313}$-$5B_{88}$). These BHK-21 cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. Cell lysates were immunoprecipitated with human patient serum JHF (A) or the following HCV-specific rabbit antisera: NS5A-specific WU123 (B and C), NS5B-specific WU115 (C), and WU113 specific for both NS5A and NS5B (D). Apparently, rabbit antiserum WU113, which was raised against a fusion protein containing the C-terminal 109 residues of NS5A and the N-terminal 203 residues of NS5B, recognizes only the NS5A region but not the NS5B sequences in $5A_{297}$-$5B_{88}$ and $5A_{313}$-$5B_{88}$. Immunoprecipitated proteins were solubilized and separated by electrophoresis on 14% (A and D) or 8% (B and C) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left. In panel A, NS4A is difficult to visualize because it contains only a single methionine residue (compared with six in NS4B) and migrates as a diffuse band on this gel system.

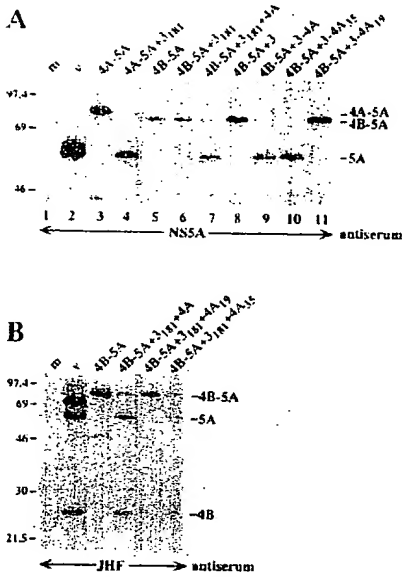


FIG. 5. Requirements for *trans*-cleavage at the 4B/5A site. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v) or vHCV1027-1207 (3_{181}). As indicated, some monolayers were also transfected with the following plasmids: pTM3/HCV1658-2420 (4A-5A), pTM3/HCV1712-2420 (4B-5A), pTM3/HCV1658-1711 (4A), pTM3/HCV1027-1657 (3), pTM3/HCV1027-1676 ($3\text{-}4A_{19}$), pTM3/HCV1027-1692 ($3\text{-}4A_{35}$), pBRTM/HCV1027-1711 (3-4A), pTM3/HCV1658-1676 ($4A_{19}$), and pTM3/HCV1658-1692 ($4A_{35}$). Cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV-specific products were immunoprecipitated with NS5A-specific antiserum WU123 (A) or human patient serum JHF (B), solubilized, and separated by 8% (A) or 10% (B) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

in the NS4A region to map NS4A sequences required for this activity. NS3-$4A_{35}$ and NS3-$4A_{19}$ contain the full-length NS3 followed by the N-terminal 35 and 19 residues of NS4A, respectively (Fig. 1). As evidenced by production of NS5A, NS3-$4A_{35}$ (Fig. 5A, lane 10), but not NS3-$4A_{19}$ (lane 11), was able to process NS4B-5A. In an earlier study, similar constructs were generated to map the location of NS4A (32). A polyprotein beginning with the C protein and extending through the N-terminal 35 residues of NS4A was efficiently processed at the 3/4A site. However, a C-terminal truncation to residue 19 of NS4A appeared to block cleavage at the 3/4A site (32). Thus, the inability of NS3-$4A_{19}$ to function for *trans*-cleavage of NS4B-5A might result from lack of cleavage at the 3/4A site and release of the NS4A N terminus rather than deletion of NS4A residues 20 to 35. To address this possibility, we examined the activity of polypeptides encompassing the N-terminal 19 and 35 residues of NS4A (called $NS4A_{19}$ and $NS4A_{35}$, respectively). $NS4A_{35}$, but not $NS4A_{19}$, was able to induce *trans*-cleavage of NS4B-5A by $NS3_{181}$ (Fig. 5B). These results indicate that the C-terminal 19 amino acids (residues 36 to 54) of NS4A, which contain 8 to 9 highly conserved, charged residues (see Discussion), are not required for *trans*-cleavage at the 4B/5A site.

Cleavage at the 3/4A and 4A/4B sites, which flank NS4A, may also require NS4A sequences for efficient cleavage (see Discussion). However, since the 5A/5B site can be efficiently cleaved in the absence of NS4A (Fig. 4C and D), this protein is not absolutely required for NS3 serine proteinase activity. For development of an in vitro proteinase assay that does not require NS4A, substrates containing the 5A/5B site should be good candidates.

***trans*-Cleavage between HCV-H and HCV-BK strains.** Viral proteinases, which are important for polyprotein processing and viral replication, present attractive targets for development of antiviral therapeutic agents. Since sequence analysis of HCV isolates has uncovered considerable genetic diversity, the success of a proteinase inhibitor strategy will depend at least in part on the conservation of proteinase-substrate interactions among different HCV types. In one classification scheme, six major genotypes or types (from 1 to 6) are distinguished, with some types further divided into related subtypes (64, 65). HCV-H (26) and HCV-BK (67) are members of the 1a and 1b subtypes, respectively, which represent the major subtypes in the United States and Japan. These two strains share 90 and 87% amino acid sequence identities in the serine proteinase domain and in the NS3-5B region, respectively. To examine the conservation or divergence of proteinase-substrate interactions among different HCV strains, we compared the ability

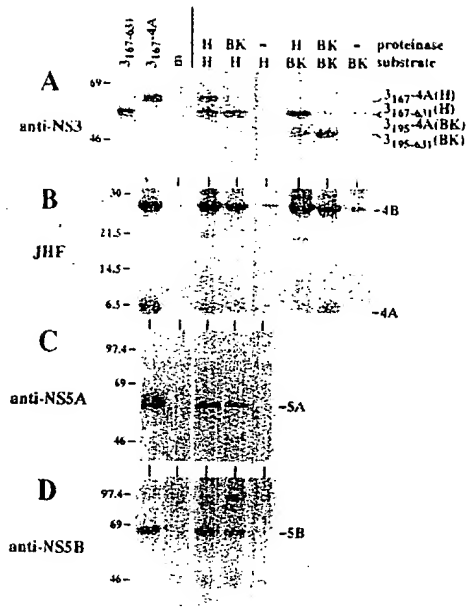


FIG. 6. *trans*-Cleavage between HCV-H and HCV-BK polypeptides. BHK-21 cell monolayers were infected with vTF7-3 alone (m) or in combination with vHCV827-3011 (v). For *trans*-cleavage experiments, the polyprotein substrates were pBRTM/HCV1193-3011 for the H strain (H) or pBRTM/HCV-BK1221-3010 for the BK strain (BK). The serine proteinase domains of both strains were expressed as the source of the proteolytic activities: vHCV1027-1207 for the H strain and pTM3/HCV-BK1027-1207 for the BK strain. The absence (−) of certain expression constructs is indicated. Cells were labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV-specific products were immunoprecipitated with the following antisera: NS3-specific WU117 (A), NS5A-specific WU123 (C), NS5B-specific WU115 (D), or human patient serum JHF (B). The immunoprecipitated proteins were solubilized and separated on 8% (A, C, and D) or 14% (B) polyacrylamide–SDS gels. HCV-specific proteins are indicated on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

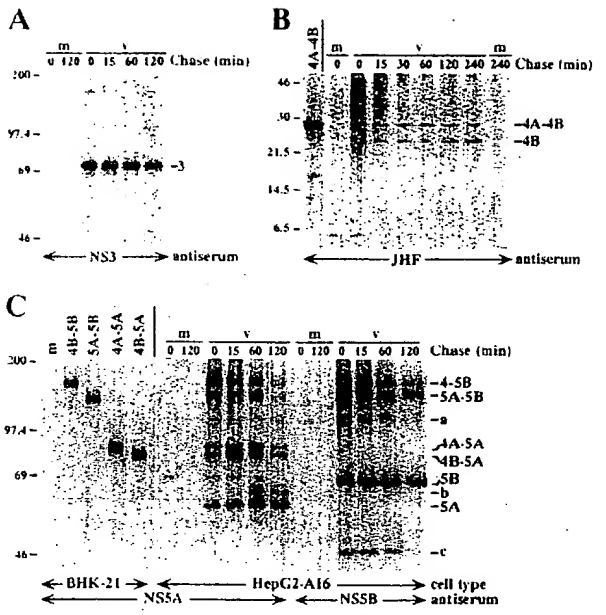


FIG. 7. Pulse-chase analysis of processing in the NS3-4-5 regions. HepG2-A16 cells were infected with vTF7-3 alone (m) or coinfected with vTF7-3 and vHCV1-3011 (v), pulse labeled with ^{35}S-protein labeling mixture for 20 min, and chased for the indicated time as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated with the following antisera: NS3-specific WU117 (A), NS5A-specific WU123, or NS5B-specific WU115 (C) or human patient serum JHF (B). Immunoprecipitated proteins were solubilized and separated by SDS-PAGE (8% polyacrylamide) (A and C) or Tricine-SDS-PAGE (14% polyacrylamide) (B). The migration pattern of HCV NS5A-specific polyprotein markers is shown on the left in panel C. BHK-21 cells previously infected with vTF7-3 were mock transfected (m) or transfected with the indicated plasmids and labeled with ^{35}S-protein labeling mixture as described in Materials and Methods. HCV-specific proteins are identified on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

of the NS3 serine proteinases of the HCV-H or the HCV-BK strain to mediate *trans*-cleavage of homologous or heterologous polyprotein substrates.

For the H strain, we used the $NS3_{181}$ proteinase and the $NS3_{167}$-5B substrate described above (Fig. 1). For the HCV-BK proteinase, we made a similar construct expressing the N-terminal 181 amino acids of HCV-BK NS3 [$NS3_{181}$(BK)]. The HCV-BK substrate was a polyprotein beginning with residue 195 of NS3 and extending through NS5B [$NS3_{195}$-5B(BK)]. When $NS3_{167}$-5B was coexpressed with $NS3_{181}$, processing at all four sites occurred, as evidenced by the appearance of $NS3_{167\text{-}631}$, NS4A, NS4B, NS5A, and NS5B (Fig. 6). For the BK strain, $NS3_{181}$(BK) was able to mediate *trans*-cleavage at the 3/4A, 4A/4B, and 4B/5A sites of $NS3_{195}$-5B(BK), as indicated by the production of $NS3_{195\text{-}631}$(BK), NS4A, and NS4B (Fig. 6A and B). Thus far, we have been unable to identify the HCV-BK NS5A and NS5B cleavage products by using HCV-H NS5A- or NS5B-specific rabbit antisera or HCV-positive patient antisera collected in the United States. As shown in Fig. 6, the serine proteinase domain of either strain was fully active at mediating *trans*-cleavage of the heterologous substrate from the other strain. $NS3_{181}$(BK) cleaved $NS3_{167}$-5B of H strain to $NS3_{167\text{-}631}$, NS4A, NS4B, NS5A, and NS5B (Fig. 6). Likewise, $NS3_{195}$-5B(BK) was processed by $NS3_{181}$ of the H strain to produce $NS3_{195\text{-}631}$(BK), NS4A, and NS4B (Fig. 6A and B). Thus, at least as assessed by this *trans*-processing assay, these two different HCV subtypes do not appear to have diverged significantly in terms of serine proteinase-substrate recognition.

Kinetics of processing in the HCV NS region. Besides defining the minimal domains required for serine proteinase activity, it is also of interest to understand the processing reactions that occur in the full-length HCV polyprotein. In other viral systems, polyprotein cleavages that occur in *cis* versus those occurring in *trans* can be important for regulating RNA replicase function. Such regulation is possible when polyproteins, processing intermediates, and mature cleavage products have distinct roles in replication (44). To begin to examine processing pathways and kinetics in the NS3-4-5 region, pulse-chase experiments were carried out in HepG2-A16 cells by using a vaccinia virus-HCV recombinant, vHCV1-3011, which expresses the entire HCV-H polyprotein (47). As shown in Fig. 7A, NS3 was readily visible after a 20-min pulse and was not associated with any higher-molecular-mass polyprotein precursors, indicating that cleavage at both the 2/3 and 3/4A sites occurs very rapidly, possibly in *cis*.

In contrast, processing in the NS4-5 region was generally

slower and a number of processing intermediates were readily identified. As shown in Fig. 7B, NS4B was readily visible after a 20-min pulse. A 29-kDa protein comigrating with the product expressed from pTM3/HCV1658-1972 (NS4A-4B) was identified as the NS4A-4B polyprotein (Fig. 7B). A decrease in the level of NS4A-4B was accompanied by an increase in the amount of NS4B, which suggests that NS4A-4B can be a precursor for NS4B and NS4A. NS4A was not observed in this experiment, probably because of its low methionine content (only one) and inefficient expression by vHCV1-3011 (32). Four predominant NS5A-containing polyproteins of 160, 135, 87, and 82 kDa were observed after the 20-min pulse (Fig. 7C). NS4-specific antiserum recognized all of these species except for the 135-kDa polyprotein (data not shown), whereas the NS5B-specific antiserum recognized only the 160- and 135-kDa polyproteins (Fig. 7C). On the basis of their apparent molecular mass, immunoreactivity, and comigration with marker polyproteins (Fig. 7C), these four polyproteins were tentatively identified as NS4-5B (160 kDa), NS5A-5B (135 kDa), NS4A-5A (87 kDa), and NS4B-5A (82 kDa). It is unclear whether the 160-kDa polyprotein NS4-5B begins with NS4A or NS4B or is a mixture of both of these species. The presence of these four polyproteins suggests that there are several alternative pathways for processing the NS4-5 region (see Discussion for more details). Over a 60-min chase, the level of NS5A (58 kDa) increased significantly and was accompanied by a decrease in the levels of NS5A-5B and NS4-5B (Fig. 7C), suggesting that these two polyproteins may be the precursors to NS5A. Because the levels of NS4B-5A and NS4A-5A increased initially, and then decreased during the chase period (Fig. 7C), they probably represent processing intermediates between NS4-5B and NS5A. An NS5A-specific protein of 62 kDa (indicated as b in Fig. 7C), barely detectable after 15 min of chase, became more apparent after 60 min. In a previous study, several minor NS5A-specific species with slower mobility were observed in addition to the dominant 58-kDa NS5A protein (32). Two additional faint bands of 107 and 47 kDa (labeled a and c, respectively, in Fig. 7C) were observed with the NS5B-specific antiserum. Product a was also recognized by NS5A-specific antiserum. These two proteins remain to be defined, but they may reflect additional proteolytic processing within the NS5B region. Although NS4-5B and NS5A-5B were likely precursors to NS5B, the level of NS5B did not change significantly over a 60-min chase period (Fig. 7C), and this protein appeared to be unstable relative to most of the other HCV-encoded proteins (see below). On the other hand, NS3 (Fig. 7A) and NS4B (Fig. 7B) were stable up to 2 h, while a slight decrease in the level of NS5A was observed (Fig. 7C).

Instability of the NS5B protein. While NS3 was very stable during prolonged chase periods, NS5A and, in particular, NS5B, the putative HCV RNA polymerase, appeared to be rather unstable. NS5A disappeared with a half-life of approximately 170 min, and NS5B disappeared with a half-life of about 70 min (data not shown). This observation is potentially interesting because some positive-strand viruses tightly regulate the level of their RNA-dependent RNA polymerase. Additionally, the p75 protein of bovine viral diarrhea virus, the HCV NS5B homolog, is unstable in bovine viral diarrhea virus-infected cells (21). To determine whether other HCV-encoded proteins might be responsible for the instability of NS5B, we expressed two different forms of HCV NS5B. One form (Met-NS5B) included the entire NS5B region preceded by two non-HCV residues, Met-Gly. A second construct encoded a ubiquitin fusion protein consisting of the 76-residue ubiquitin monomer fused in frame to the N terminus of NS5B (Ubi-NS5B). Cleavage of this ubiquitin fusion protein by

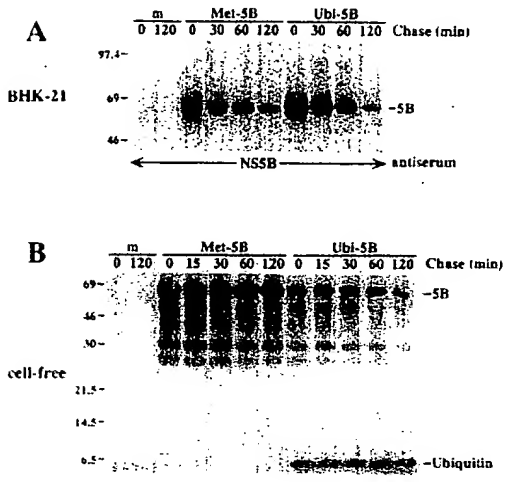


FIG. 8. Stability of NS5B expressed in BHK-21 cells or by cell-free translation. (A) BHK-21 cells previously infected with vTF7-3 were mock transfected (m) or transfected with one of the following plasmids: pTM3/HCV2421-3011 (Met-5B) or pTM3/Ubiquitin-HCV2421-3011 (Ubi-5B). The cell monolayers were pulse-labeled with ^{35}S-protein labeling mixture for 20 min and chased for the indicated times as described in Materials and Methods. Cell lysates were prepared, and HCV NS5B-specific products were immunoprecipitated by the rabbit antiserum WU115. Immunoprecipitated proteins were solubilized and separated by SDS-PAGE (8% polyacrylamide). HCV-specific proteins are identified on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) from Amersham are indicated on the left. (B) Translations with RNA transcripts from pTM3/HCV2421-3011 or pTM3/Ubiquitin-HCV2421-3011 or without any transcript (m) were incubated for 60 min at 30°C in reticulocyte lysate in the presence of [^{35}S]methionine. Translation reactions were terminated by the addition of RNase A, cycloheximide, and excess cold methionine and then were chased for the indicated times. The translation products were solubilized and analyzed by SDS-PAGE (14% polyacrylamide). The identities of proteins are shown on the right, and the sizes of ^{14}C-labeled protein molecular mass markers (in kilodaltons) are indicated on the left.

cellular ubiquitin carboxy-terminal hydrolase should produce NS5B with its authentic N-terminal Ser residue (4). As shown in Fig. 8A, the Ubi-NS5B fusion protein was completely processed to NS5B after a 20-min pulse of transfected BHK-21 cells. Both forms of the NS5B proteins were unstable, as evidenced by the rapid decline in the level of NS5B (Fig. 8A). The approximate half-lives were 90 min for Met-NS5B and 70 min for NS5B produced by cleavage of Ubi-NS5B. Coexpression of $NS3_{181}$ had no significant effect on the stability of NS5B (data not shown). These results indicate that the instability of NS5B is not due to the presence of other HCV-encoded proteinases or proteins. Rather, NS5B is inherently unstable and is probably degraded through a cellular pathway.

In an attempt to devise an in vitro assay to study NS5B degradation, we examined the stability of Met-NS5B or Ubi-NS5B produced by cell-free translation of RNA transcripts in rabbit reticulocyte lysates (Fig. 8B). Translation reactions were terminated by addition of RNase A, cycloheximide, and excess cold methionine and were chased for the indicated periods. Control experiments showed no further incorporation of [^{35}S]methionine after the addition of these three reagents (data not shown). Ubi-NS5B fusion proteins were completely processed to NS5B and ubiquitin (predicted molecular mass of 8.6 kDa) after a 60-min incubation. Although slight decreases

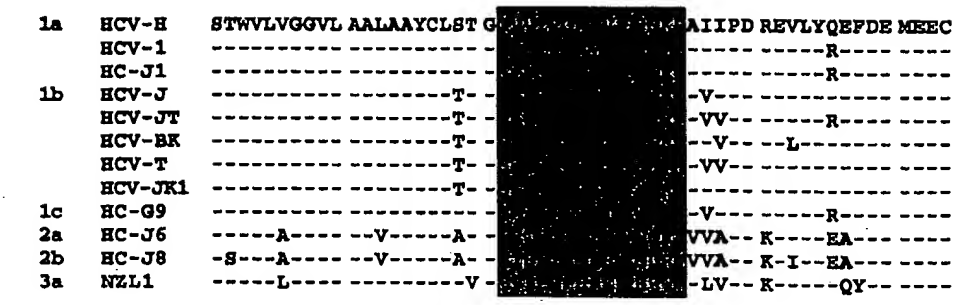


FIG. 9. Alignment of NS4A sequences. The predicted NS4A amino acid sequences are aligned for selected HCV isolates from six subtypes (indicated on the left): HCV-H (38), HCV-1 (18), HC-J1 (accession no. D10749), HCV-J (39), HCV-JT (68), HCV-BK (67), HCV-T (16), HCV-JK1 (accession no. S18030), HC-G9 (53), HC-J6 (55), HC-J8 (54), and HCV-NZL1 (60). The single-letter code for amino acids is used. Hyphens indicate residues identical to those of the HCV-H strain sequence. The 14-residue segment (residues 22 to 35) implicated in *trans*-cleavage at the 4B/5A site is shaded. Accession numbers for unpublished sequences are given above in parentheses.

in the levels of the proteins were apparent over the 2-h chase, both Met-NS5B and NS5B produced by cleavage of Ubi-NS5B were quite stable in the reticulocyte lysates, making it difficult to assess the role of the ubiquitin-mediated degradation pathway (which is present in reticulocyte lysates [19]) in the turnover of NS5B.

DISCUSSION

It has been previously shown that an active NS3 serine proteinase is required for processing at four cleavage sites in the HCV NS3-4-5 region. The results presented here clearly demonstrate that the proteinase domain, expressed as a 181-residue N-terminal fragment of NS3, is able to mediate *trans*-cleavage at all four sites. Bartenschlager et al. (6) recently reported similar results showing that a fragment of the polyprotein, including the 212 N-terminal residues of NS3 and a 20-residue extension into the NS2 region, could also mediate *trans*-cleavage at the 4A/4B, 4B/5A, and 5A/5B sites. Our results, as well as those of two recent studies (6, 23), indicate that NS4A is absolutely required for the 4B/5A cleavage. Failla et al. (23) also showed that NS4A of HCV-BK, supplied in *trans*, was required for cleavage at the 3/4A and 4B/5A sites and improves the efficiency of processing at the 4A/4B and 5A/5B sites. On the basis of these results, it was suggested that NS4A functions as a general effector or cofactor for NS3 serine proteinase-mediated cleavage in the NS3-4-5 region. Virus-encoded cofactors required for serine proteinase activity have also been found for other members of the family *Flaviviridae*. The most dramatic example is the NS2B protein of flaviviruses, which is absolutely required for NS3 serine proteinase-mediated cleavage at all structural and nonstructural dibasic sites (2, 11, 15, 24, 45, 48, 56, 57, 70, 72). As discussed by Failla et al. (23), the pestivirus p10 protein may be the functional homolog of HCV NS4A, because sequences in this region of the pestivirus polyprotein appear to be required for the serine proteinase-dependent cleavage between p58 and p75 (the two C-terminal products of the pestivirus polyprotein possibly equivalent to HCV NS5A and NS5B, respectively) (71). For HCV, NS4A is required for only three cleavages mediated by the serine proteinase (3/4A, 4A/4B, and 4B/5A). While Failla et al. showed that NS4A can increase *trans*-cleavage efficiency at the 5A/5B site (23), we found that certain substrates containing this site could be processed efficiently in the absence of NS4A. Hence, the HCV serine proteinase-dependent cleavages can be separated into at least two types: (i) cleavages at the 3/4A, 4A/4B, and 4B/5A sites, which are located adjacent to hydrophobic sequences and require NS4A as a cofactor; and (ii) cleavage at the 5A/5B site, which can occur in the absence of NS4A.

Although the mechanism(s) by which NS4A functions in proteolytic processing at type 1 sites remains to be determined, several possibilities can be envisioned. (i) NS4A may act as a molecular chaperone to facilitate folding of the serine proteinase domain into an active enzyme. If the active form of the proteinase is the same for cleavage at both type 1 and type 2 sites, then this model implies that type 1 substrates are suboptimal and require higher concentrations of active proteinase for efficient *trans*-cleavage. (ii) NS4A may bind to type 1 substrates, the proteinase domain, or both to facilitate proteinase-substrate interactions and cleavage. (iii) NS4A may facilitate proteolysis of membrane-associated type 1 substrates by interacting with the proteinase domain and localizing it to the membrane compartment. Given that NS4A is required for cleavage at three different sites, it is tempting to propose that it functions via direct interaction with the proteinase domain. Thus far, unlike the flavivirus proteinase, which consists of a stable complex of NS2B and NS3 (3, 14), there is no direct evidence for association between the HCV NS3 and NS4A proteins. Suggestive evidence has been obtained, however, from in vitro studies in which NS3 was found to become membrane associated when the cell-free translation product included the NS4A region (35).

Although NS4A is only 54 residues in length, we showed that a fragment of only 35 N-terminal residues, coexpressed with the serine proteinase domain, was sufficient for *trans*-cleavage at the 4B/5A site. Failla et al. (23) reported that a polypeptide consisting of the C-terminal 33 residues of NS4A and NS4B facilitated *trans*-cleavage at the 4B/5A site. Although flanking sequences may contribute to NS4A activity, these data suggest that a 14-residue segment (residues 22 to 35) of NS4A may be critical for cleavage at the 4B/5A site. As shown in Fig. 9, the HCV NS4A protein sequence is highly conserved among HCV strains and consists of a hydrophobic N-terminal portion, a central region implicated in *trans*-cleavage at the 4B/5A site (highlighted in Fig. 9), and a highly charged acidic C-terminal segment. Although somewhat less conserved than other regions of NS4A (especially in comparison with HCV-J6 and HCV-J8), this central region contains two positively-charged residues, several hydrophobic amino acids, and an absolutely conserved Gly at position 27. The importance of these residues for NS4A *trans*-cleavage activity is currently being tested by site-directed mutagenesis.

The lack of *trans*-cleavage at the 3/4A site in previous studies led to the suggestion that cleavage at this site occurred in *cis* (5, 69). This cleavage has recently been shown to be insensitive to dilution, providing direct evidence for a *cis* mechanism (6). These observations are consistent with the results of pulse-chase analyses in which we (this report) and others (6) were unable to detect NS3-related precursors. Thus, in the current model, both the 2/3 and 3/4A cleavages are catalyzed by two distinct viral proteinases in *cis*. Of interest is the observation that substrates with an inactivated serine proteinase domain were resistant to *trans*-cleavage at the 3/4A site. Efficient *trans*-cleavage was observed, however, when the inactivated proteinase domain was deleted. Although other possibilities exist, these results, together with the observation that NS4A sequences are required for cleavage at the 3/4A site (23), suggest that during translation of the polyprotein, the serine proteinase domain interacts with nascent NS4A to assume a conformation capable of *cis*-cleavage at the 3/4A site. In the case of the substrate with the inactivated proteinase, this intermediate still forms but is frozen because it is inactive for *cis*-cleavage. Thus, the 3/4A site of this substrate is not accessible to *trans*-acting proteinase, because it is probably bound in the substrate binding pocket of the inactive autoproteinase.

In contrast to the rapid cleavages observed at the 2/3 and 3/4A sites, processing at the 4A/4B, 4B/5A, and 5A/5B sites was slower and appeared to involve multiple pathways (this study and reference 6). An obligate processing order was not observed, which is consistent with results from a study in which mutations blocking cleavage at each of these three sites had no significant effect on processing at other sites (40). Similar results have been obtained for flaviviruses (45, 46, 52, 56). It is important to emphasize that the processing pathways and kinetics observed in mammalian transient expression assays may not accurately reflect the situation in HCV-infected cells. In particular, *trans*-processing reactions, which are important for temporal regulation of RNA synthesis for other viruses (for example, see reference 44), would be expected to be sensitive to the concentration of *trans*-acting factors, which may be much lower in HCV-infected cells. Hence, these issues should be reexamined when systems become available for studying HCV replication in cell cultures.

Using both the vaccinia virus-T7 and the Sindbis virus replicon expression systems, we found that the NS5B protein was unstable compared with the other polyprotein cleavage products (8) (Fig. 7 and 8 and data not shown). Turnover of NS5B was similar whether the protein was expressed as part of the full-length polyprotein or independently as a ubiquitin fusion protein. In contrast to the results in cell culture assays, NS5B was found to be relatively stable in reticulocyte lysates. Since NS5B is the putative HCV RNA-dependent RNA polymerase (51), down-regulation of this protein could play an important regulatory role in virus replication, as has been found for the RNA polymerase of alphaviruses (see reference 66 for a review). For the pestivirus bovine viral diarrhea virus, the putative RNA-dependent RNA polymerase (p75) is unstable, with a half-life of less than 60 min in bovine viral diarrhea virus-infected cells (21). However, the NS5 protein of flaviviruses, which is not cleaved into two proteins, is relatively stable (13). In contrast to our results, Bartenschlager et al. found NS5B (of a strain similar to HCV-J) to be quite stable when expressed in HeLa cells with a vaccinia virus recombinant (6). The reason for the discrepancy is unclear, but it could reflect a difference in the sequence of the expressed NS5B protein or in the cells used for the expression studies. As mentioned above, these issues need to be reexamined in a system that supports HCV RNA replication.

Finally, there is considerable interest in the HCV proteinases as targets for development of new antivirus therapies. The general usefulness of such compounds will depend in part on their ability to inhibit the proteinases of diverse HCV types. Although it will be important to test more divergent proteinase-substrate combinations, the ability of the HCV-H serine proteinase (subtype 1a) to *trans*-process an HCV-BK substrate (subtype 1b), and vice versa, suggests that the essential elements of recognition may be conserved. This is encouraging for the development of broadly effective serine proteinase inhibitors.

ACKNOWLEDGMENTS

We thank Bernie Moss for providing pTM3 and vTF7-3; Hiroto Okayama and Akihisa Takamizawa for HCV-BK cDNA clones; Henry Hsu and Harry Greenberg for human patient serum JHF; and Sean Amberg, Julie Lemm, Sasha Kolykhalov, and Karen Reed for critical reading of the manuscript.

This work was supported by a grant from the Public Health Service (CA57973). C.L. is a predoctoral candidate and was supported in part by the Division of Biology and Biomedical Sciences at Washington University. B.M.P. was supported in part by the Ministry of Welfare, T076-ETT (90-93), Hungary.

REFERENCES

1. **Alter, M. J., S. C. Hadler, F. N. Judson, A. Mares, W. J. Alexander, P. Y. Hu, J. K. Miller, L. A. Moyer, H. A. Fields, D. W. Bradley, and H. S. Margolis.** 1990. Risk factors for acute non-A, non-B hepatitis in the United States and association with hepatitis C virus infection. JAMA **264**:2231–2235.
2. **Amberg, S. M., A. Nestorowicz, D. W. McCourt, and C. M. Rice.** 1994. NS2B-3 proteinase-mediated processing in the yellow fever virus structural region: in vitro and in vivo studies. J. Virol. **68**:3794–3802.
3. **Arias, C. F., F. Preugschat, and J. H. Strauss.** 1993. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. Virology **193**:888–899.
4. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science **234**:179–186.
5. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1993. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J. Virol. **67**:3835–3844.
6. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1994. Kinetic and structural analyses of hepatitis C virus polyprotein processing. J. Virol. **68**:5045–5055.
7. **Bazan, J. F., and R. J. Fletterick.** 1990. Structural and catalytic models of trypsin-like viral proteases. Semin. Virol. **1**:311–322.
8. **Bredenbeek, P. J., I. Frolov, C. M. Rice, and S. Schlesinger.** 1993. Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. J. Virol. **67**:6439–6446.
9. **Brockman, W. W., and D. Nathans.** 1974. The isolation of simian virus 40 variants with specifically altered genomes. Proc. Natl. Acad. Sci. USA **71**:942–946.
10. **Bukh, J., R. H. Purcell, and R. H. Miller.** 1993. At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc. Natl. Acad. Sci. USA **90**:8234–8238.
11. **Chambers, T. J., A. Grakoui, and C. M. Rice.** 1991. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J. Virol. **65**:6042–6050.
12. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. **44**:649–688.
13. **Chambers, T. J., D. W. McCourt, and C. M. Rice.** 1990. Production of yellow fever virus proteins in infected cells: identification of discrete polyprotein species and analysis of cleavage kinetics using

region-specific polyclonal antisera. Virology 177:159–174.
14. Chambers, T. J., A. Nestorowicz, S. M. Amberg, and C. M. Rice. 1993. Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. J. Virol. 67:6797–6807.
15. Chambers, T. J., R. C. Weir, A. Grakoui, D. W. McCourt, J. F. Bazan, R. J. Fletterick, and C. M. Rice. 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA 87:8898–8902.
16. Chen, P.-J., M.-H. Lin, K.-F. Tai, P.-C. Liu, C.-J. Lin, and D.-S. Chen. 1992. The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5′ termini of viral genomic and antigenomic RNA. Virology 188:102–113.
17. Choo, Q.-L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244:359–362.
18. Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton. 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA 88:2451–2455.
19. Ciechanover, A., J. A. DiGiuseppe, B. Bercovich, A. Orian, J. D. Richter, A. L. Schwartz, and G. M. Brodeur. 1991. Degradation of nuclear oncoproteins by the ubiquitin system *in vitro*. Proc. Natl. Acad. Sci. USA 88:139–143.
20. Collett, M. S. 1992. Molecular genetics of pestiviruses. Comp. Immunol. Microbiol. Infect. Dis. 15:145–154.
21. Collett, M. S., M. A. Wiskerchen, E. Welniak, and S. K. Belzer. 1991. Bovine viral diarrhea virus genomic organization. Arch. Virol. 1991(Suppl. 3):19–27.
22. Eckart, M. R., M. Selby, F. Masiarz, C. Lee, K. Berger, K. Crawford, C. Kuo, G. Kuo, M. Houghton, and Q.-L. Choo. 1993. The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. Biochem. Biophys. Res. Commun. 192: 399–406.
23. Failla, C., L. Tomei, and R. De Francesco. 1994. Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. J. Virol. 68:3753–3760.
24. Falgout, B., M. Pethel, Y.-M. Zhang, and C.-J. Lai. 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65:2467–2475.
25. Falkner, F. G., and B. Moss. 1988. *Escherichia coli gpt* gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. 62:1849–1854.
26. Feinstone, S. M., H. J. Alter, H. P. Dienes, Y. Shimizu, H. Popper, D. Blackmore, D. Sly, W. T. London, and R. H. Purcell. 1981. Non-A, non-B hepatitis in chimpanzees and marmosets. J. Infect. Dis. 144:588–598.
27. Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown. 1991. Classification and nomenclature of viruses: fifth report of the International Committee on Taxonomy of Viruses. Arch. Virol. 1991(Suppl. 2):223.
28. Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss. 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 83:8122–8126.
29. Gorbalenya, A. E., A. P. Donchenko, E. V. Koonin, and V. M. Blinov. 1989. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. Nucleic Acids Res. 17:3889–3897.
30. Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice. 1993. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J. Virol. 67:2832–2843.
31. Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice. 1993. A second hepatitis C virus-encoded proteinase. Proc. Natl. Acad. Sci. USA 90:10583–10587.
32. Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone, and C. M. Rice. 1993. Expression and identification of hepatitis C virus polyprotein cleavage products. J. Virol. 67:1385–1395.
33. Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa, and K. Shimotohno. 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc. Natl. Acad. Sci. USA 88:5547–5551.
34. Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno. 1993. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J. Virol. 67:4665–4675.
35. Hijikata, M., H. Mizushima, Y. Tanji, Y. Komoda, Y. Hirowatari, T. Akagi, N. Kato, K. Kimura, and K. Shimotohno. 1993. Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. Proc. Natl. Acad. Sci. USA 90:10773–10777.
36. Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo. 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology 14:381–388.
37. Hruby, D. E., L. A. Guarino, and J. R. Kates. 1979. Vaccinia virus replication. I. Requirement for the host-cell nucleus. J. Virol. 29:705–715.
38. Inchauspe, G., S. Zebedee, D.-H. Lee, M. Sugitani, M. Nasoff, and A. M. Prince. 1991. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. Proc. Natl. Acad. Sci. USA 88:10292–10296.
39. Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno. 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA 87:9524–9528.
40. Kolykhalov, A. A., E. V. Agapov, and C. M. Rice. 1994. Specificity of the hepatitis C virus NS3 serine protease: effects of substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. J. Virol. 68:7525–7533.
41. Kuo, G., Q.-L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redeker, R. H. Purcell, T. Miyamura, J. L. Dienstag, M. J. Alter, C. E. Stevens, G. E. Tegtmeier, F. Bonino, M. Colombo, W.-S. Lee, C. Kuo, K. Berger, J. R. Shuster, L. R. Overby, D. W. Bradley, and M. Houghton. 1989. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science 244:362–364.
42. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
43. Laskey, R. A., and A. D. Mills. 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. 56:335–341.
44. Lemm, J. A., T. Rümenapf, E. G. Strauss, J. H. Strauss, and C. M. Rice. 1994. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus and plus-strand RNA synthesis. EMBO J. 13:2925–2934.
45. Lin, C., S. M. Amberg, T. J. Chambers, and C. M. Rice. 1993. Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. J. Virol. 67:2327–2335.
46. Lin, C., T. J. Chambers, and C. M. Rice. 1993. Mutagenesis of conserved residues at the yellow fever virus 3/4A and 4B/5 dibasic cleavage sites: effects on cleavage efficiency and polyprotein processing. Virology 192:596–604.
47. Lin, C., B. D. Lindenbach, B. M. Prágai, D. W. McCourt, and C. M. Rice. 1994. Processing of the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. J. Virol. 68:5063–5073.
48. Lobigs, M. 1992. Proteolytic processing of a Murray Valley encephalitis virus nonstructural polyprotein segment containing the viral proteinase: accumulation of a NS3-4A precursor which requires mature NS3 for efficient processing. J. Gen. Virol. 73:2305–2312.
49. Mackett, M., and G. L. Smith. 1986. Vaccinia virus expression vectors. J. Gen. Virol. 67:2067–2082.
50. Manabe, S., I. Fuke, O. Tanishita, C. Kaji, Y. Gomi, S. Yoshida, C. Mori, A. Takamizawa, I. Yoshida, and H. Okayama. 1994. Pro-

duction of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. Virology 198:636–644.
51. **Matsuura, Y., and T. Miyamura.** 1993. The molecular biology of hepatitis C virus. Semin. Virol. 4:297–304.
52. **Nestorowicz, A., T. J. Chambers, and C. M. Rice.** 1994. Mutagenesis of the yellow fever virus NS2A/2B cleavage site: effects on proteolytic processing, viral replication and evidence for alternative processing of the NS2A protein. Virology 199:114–123.
53. **Okamoto, H., M. Kojima, M. Sakamoto, H. Iizuka, S. Hadiwandowo, S. Suwignyo, Y. Miyakawa, and M. Mayumi.** 1994. The entire nucleotide sequence and classification of a hepatitis C virus isolate of a novel genotype from an Indonesian patient with chronic liver disease. J. Gen. Virol. 75:629–635.
54. **Okamoto, H., K. Kurai, S.-I. Okada, K. Yamamoto, H. Iizuka, T. Tanaka, S. Fukuda, F. Tsuda, and S. Mishiro.** 1992. Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology 188:331–341.
55. **Okamoto, H., S. Okada, Y. Sugiyama, K. Kurai, H. Iizuka, A. Machida, Y. Miyakawa, and M. Mayumi.** 1991. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J. Gen. Virol. 72:2697–2704.
56. **Preugschat, F., and J. H. Strauss.** 1991. Processing of nonstructural proteins NS4A and NS4B of dengue 2 virus *in vitro* and *in vivo*. Virology 185:689–697.
57. **Preugschat, F., C.-W. Yao, and J. H. Strauss.** 1990. In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J. Virol. 64:4364–4374.
58. **Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang.** 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. J. Virol. 61:3809–3819.
59. **Rice, C. M., and J. H. Strauss.** 1982. Association of Sindbis virion glycoproteins and their precursors. J. Mol. Biol. 154:325–348.
60. **Sakamoto, M., Y. Akahane, F. Tsuda, T. Tanaka, D. G. Woodfield, and H. Okamoto.** 1994. Entire nucleotide sequence and characterization of a hepatitis C virus of genotype V/3a. J. Gen. Virol. 75:1761–1768.
61. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
62. **Schägger, H., and G. von Jagow.** 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range of 1 to 100 kDa. Anal. Biochem. 166:368–379.
63. **Shimotohno, K.** 1993. Hepatocellular carcinoma in Japan and its linkage to infection with hepatitis C virus. Semin. Virol. 4:305–312.
64. **Simmonds, P., E. C. Holmes, T.-A. Cha, S.-W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea.** 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS5 region. J. Gen. Virol. 74:2391–2399.
65. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 1994. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS5 regions. J. Gen. Virol. 75:1053–1061.
66. **Strauss, J. H., and E. G. Strauss.** 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58:491–562.
67. **Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama.** 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. 65:1105–1113.
68. **Tanaka, T., N. Kato, M. Nakagawa, Y. Ootsuyama, M.-J. Cho, T. Nakazawa, M. Hijikata, Y. Ishimura, and K. Shimotohno.** 1992. Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals. Virus Res. 23:39–53.
69. **Tomei, L., C. Failla, E. Santolini, R. De Francesco, and N. La Monica.** 1993. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. J. Virol. 67:4017–4026.
70. **Wengler, G., G. Czaya, P. M. Färber, and J. H. Hegemann.** 1991. *In vitro* synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. J. Gen. Virol. 72:851–858.
71. **Wiskerchen, M., and M. S. Collett.** 1991. Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. Virology 184:341–350.
72. **Zhang, L., P. M. Mohan, and R. Padmanabhan.** 1992. Processing and localization of dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. J. Virol. 66:7549–7554.

JOURNAL OF VIROLOGY, Mar. 1986, p. 826–832
0022-538X/86/030826-07$02.00/0

Vol. 57, No. 3

Bovine Leukemia Virus Protease: Purification, Chemical Analysis, and In Vitro Processing of *gag* Precursor Polyproteins

YOSHIYUKI YOSHINAKA,†* IYOKO KATOH,† TERRY D. COPELAND, GARY W. SMYTHERS, AND STEPHEN OROSZLAN‡

Laboratory of Molecular Virology and Carcinogenesis, Litton Bionetics, Inc., Basic Research Program, National Cancer Institute Frederick Cancer Research Facility, Frederick, Maryland 21701

Received 26 July 1985/Accepted 19 November 1985

Bovine leukemia virus protease was purified to homogeneity and assayed by using murine leukemia virus $Pr65^{gag}$, a polyprotein precursor of the viral core structural proteins, as the substrate. A chemical analysis of the protease, including an amino acid composition and NH_2- and COOH-terminal amino acid sequence analysis, revealed that it has an M_r of 14,000 and is encoded by a segment of the viral RNA located between the *gag* gene and the putative reverse transcriptase gene. As expected from the nucleotide sequence data (Rice et al., Virology 142:357–377, 1985), the reading frame for the protease is different from both the *gag* and reverse transcriptase reading frames. The 5′ end of the protease open reading frame extends 38 codons upstream from the codon for the NH_2-terminal residue of the mature viral protease and overlaps the *gag* open reading frame by 7 codons. The 3′ end of the protease open reading frame extends 26 codons beyond the codon for the COOH-terminal residue of the mature protease and overlaps 8 codons of the reverse transcriptase open reading frame. Several lines of evidence, such as protein mapping of the *gag* polyprotein precursor, the characteristic structure of the mRNA, and promotion of the synthesis of a *gag* polyprotein precursor by lysine tRNA in vitro, suggest that the protease could be translated by frameshift suppression of the *gag* termination codon. In vitro synthesized bovine leukemia virus *gag*-related polyproteins were cleaved by the protease into fragments which were the same size as the known components of bovine leukemia virus, suggesting that the specificity of cleavage catalyzed in vitro by the purified protease is the same as the specificity of cleavage found in the virus.

Bovine leukemia virus (BLV) is the etiologic agent of enzootic bovine leukemia (2). Structural analyses of the major viral core protein have allowed this virus, along with the recently discovered human T-cell leukemia virus (17, 26), to be classified as members of a new family of retroviruses (16). Chemical analyses of the viral structural proteins (4, 13, 14, 21) and nucleotide sequence analyses of the proviral DNAs (18–20) have shown that the gene order in the BLV genome is 5′-*gag-pol-env*-x-3′. The *gag* gene codes for core structural proteins p15, p24, and p12; the *pol* gene codes for reverse transcriptase; and the *env* gene codes for envelope proteins gp60 and p30. In addition to the structural proteins, several open reading frames have been observed, including two open reading frames in the *pol* gene and, as in human T-cell leukemia virus (23), two open reading frames downstream from the *env* gene (X region or long open reading frame region) (19, 20). Between the 3′ end of the *gag* gene and the 5′ end of the putative reverse transcriptase gene, the amino acid sequence deduced from the DNA sequence shows clear homology to the murine and avian viral proteases. Recently, we identified Moloney murine leukemia virus protease and reported that this protease is encoded by the *gag-pol* gene and is synthesized through suppression of the *gag* termination codon (UAG) (28). In this paper we describe the purification, chemical analysis, and biological specificity of the BLV protease. The possible mechanism of synthesis of the protease, which could be translated by frameshift suppression of the *gag* termination codon, is discussed.

* Corresponding author.
† Present address: Japan Immuno Research Laboratories, 17-5 Sakae, Takasaki-City, Gunma Prefecture 370, Japan.
‡ To whom requests for reprints should be sent.

MATERIALS AND METHODS

Viruses. BLV was grown in fetal lamb kidney cells (24). Gazdar murine sarcoma virus (Gz-MSV) was grown in HTG-2 cells (6). The viruses were purified by sucrose density gradient centrifugation by the Biological Products Laboratory, Program Resources, Inc., National Cancer Institute Frederick Cancer Research Facility, Frederick, Md.

Assay of protease. Gz-MSV, which itself has no protease activity and contains uncleaved $Pr65^{gag}$ (30), a homolog of the polyprotein precursor of the *gag* (group-specific antigen) gene products in other murine viruses, was used as a substrate in assays for the protease activity. Protease activity was assayed as previously described (29). In some cases, immune precipitates of BLV *gag* polyproteins synthesized in an in vitro translation system were incubated with the protease fractions under the same conditions.

Purification of the protease. The method used to purify the protease was essentially the same method used previously for the mouse protease (28). To 60 mg of purified BLV suspended in 2 ml of 0.13 M NaCl–0.01 M Tris hydrochloride (pH 7.2)–0.001 M EDTA (STE buffer), 20 volumes of cold (−70°C) acetone was added. The precipitate was dried under reduced pressure. The resulting acetone powder was then extracted with 4 ml of 0.02 M Tris hydrochloride (pH 7.2)–5 mM dithiothreitol (Sigma Chemical Co., St. Louis, Mo.) (TD buffer) containing 1.0 M NaCl at 4°C for 30 min with constant stirring, and the preparation was centrifuged at 10,000 × *g* for 10 min at 4°C. The supernatant, which contained the protease activity, was then fractionated on a Sephacryl

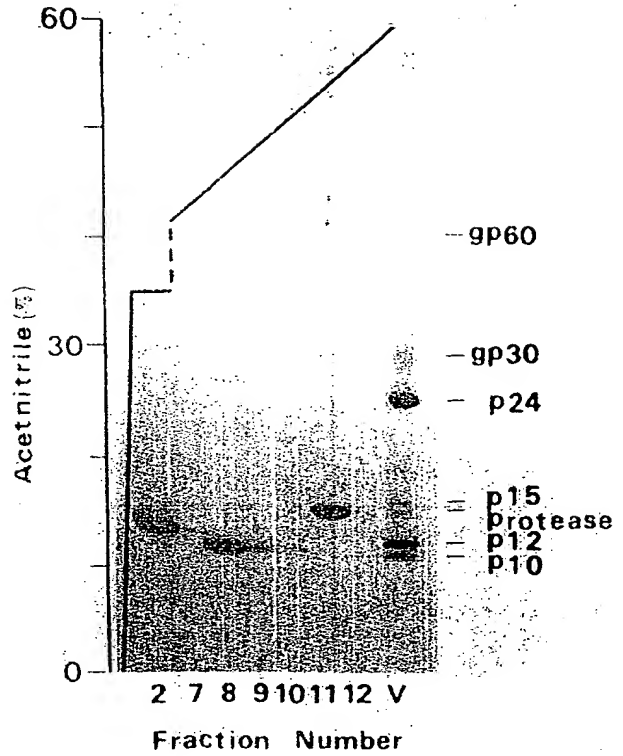


FIG. 1. Purification of protease by RP-HPLC. The protease-active fraction from Sephacryl S-200 chromatography (see text) was further fractionated by RP-HPLC on a μBondapak C_{18} column (0.39 by 30 cm; Waters Associates). The gradient conditions used were as follows: 0 to 35% acetonitrile over a period of 2 min; isocratic at 35% acetonitrile for 10 min; and 35 to 60% acetonitrile over a period of 50 min at a constant flow rate of 1.0 ml/min. One-tenth of each fraction was lyophilized and analyzed for protein composition. Staining was with Coomassie brilliant blue R-250. Lane V contained BLV (40 μg).

S-200 chromatography column (2.5 by 90 cm). After the protease activity of each fraction (3 ml) was determined, fractions were pooled and subjected to reverse-phase high-pressure liquid chromatography (RP-HPLC) by using a μBondapak C_{18} column (Waters Associates, Inc., Milford, Mass.). Fractions (5 ml) were collected, and portions were removed for protein composition analysis, sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE), and protease activity measurements.

SDS-PAGE. Various protein materials were analyzed by discontinuous SDS-PAGE (11). Proteins were visualized by staining with either Coomassie brilliant blue R-250 (Bio-Rad Laboratories, Richmond, Calif.) or silver (25). Radiofluorography was done by using a sodium salicylate treatment (3).

Amino acid composition. Samples for amino acid analysis were hydrolyzed for 24 h in vacuo with 6 N HCl containing 0.1% liquid phenol and then dried by vacuum desiccation. Analyses were performed with a Durrum model D-500 amino acid analyzer by using ninhydrin to detect the eluted amino acids.

NH_2-terminal microsequence analysis. The purified protease was subjected to automated Edman degradation in a gas phase sequenator (9) by using the program supplied by the manufacturer (Applied Biosystems, Inc.). The amino acid anilinothiazolinones were converted to phenylthiohydantoin derivatives by using 25% trifluoroacetic acid in water. Phenylthiohydantoin amino acids were identified and quantitated by RP-HPLC (8).

COOH-terminal sequence analysis. Protein samples were digested with carboxypeptidase Y for various times as previously described (14), and the amino acids released were quantitated by using the Durrum model D-500 analyzer.

In vitro translation of BLV RNA. A 20-mg portion of purified virus in STE buffer was disrupted with 1% SDS and extracted twice with water-saturated phenol. The RNA was precipitated by adding two volumes of cold (−20°C) ethanol and was collected by low-speed centrifugation. The RNA was then suspended in 0.5 ml of STE buffer containing 1% SDS and was fractionated by centrifugation on a 5 to 20% (wt/vol) sucrose density gradient (made up in STE buffer containing 0.1% SDS) at 30,000 rpm for 150 min in an SW40.1 rotor at 20°C. Fractions were collected, and the RNA was again precipitated by adding 2 volumes of cold (−20°C) ethanol. The fractions were kept overnight at −20°C, and precipitates were collected by centrifugation at 4°C. The precipitates were solubilized in 100μl of water and heated at 90°C for 2 min. A 5-μl portion of each fraction was analyzed to determine the size of the RNA species on a 1% agarose gel (12) by using bovine liver RNA as a marker. The in vitro translation experiments were performed in 25-μl incubation mixtures at 37°C for 60 min by using an in vitro translation kit (New England Nuclear Corp., Boston, Mass.)

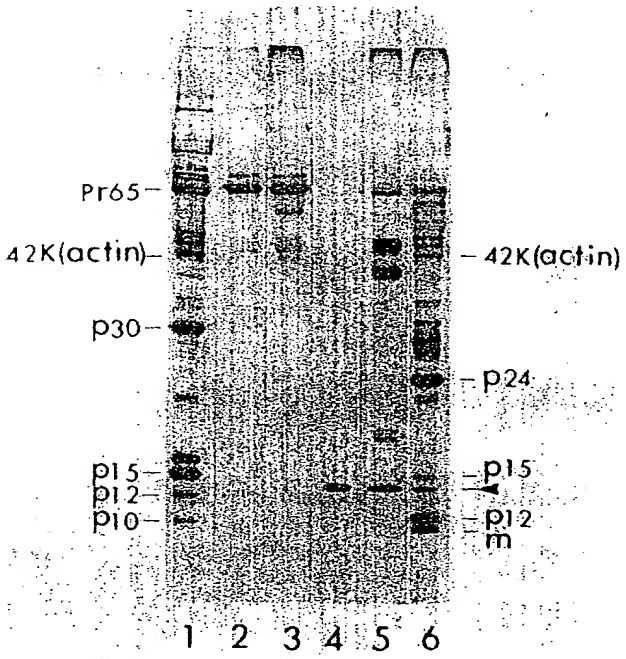


FIG. 2. Cleavage of Gz-MSV $Pr65^{gag}$ with purified BLV protease (fraction 11 in Fig. 1). Portions (2.5%) of fraction 11 from the experiment shown in Fig. 1 were lyophilized, solubilized in TD buffer (pH 7.0) containing 0.5% Nonidet P-40 (100 μl), and then incubated in the presence or absence of Gz-MSV (1.5 μg) as the substrate. After incubation at 20°C for 16 h, a 25-μl portion of each sample was analyzed by SDS-PAGE. The gel was stained with silver nitrate. Lane 1, Moloney murine leukemia virus (7 μg); lane 2, Gz-MSV (1.5 μg) alone before incubation; lane 3, Gz-MSV (1.5 μg) alone after incubation; lane 4, fraction 11 alone (0.3 μg); lane 5, fraction 11 (0.3 μg) plus Gz-MSV (1.5 μg) after incubation; lane 6, BLV (7 μg). The gel was stained with silver nitrate (25).

with 25 μCi of L-[^{35}S]methionine (New England Nuclear Corp.), a reduced concentration of potassium acetate (75 mM), 1.0 mM magnesium acetate, and RNA which was ice quenched following heating at 90°C for 2 min. Samples (1 μl) were removed to determine the radioactivity incorporated into the proteins, and the samples were analyzed by SDS-PAGE.

Immunoprecipitation. Immunoprecipitation was done as follows. Protein A-Sepharose (15 mg; Pharmacia Fine Chemicals, Piscataway, N.J.) in 0.5 ml of Dulbecco phosphate-buffered saline and 10 μl of serum were incubated for 1 h at 4°C. After washing with phosphate-buffered saline twice, the beads were incubated with a translation mixture, which was diluted 10-fold with immunoprecipitation buffer (0.02 M Tris hydrochloride pH 7.5, 0.1 M NaCl, 1% Nonidet P-40), for 16 h at 4°C. Then the beads were washed three times with immunoprecipitation buffer and once with water and analyzed by SDS-PAGE.

RESULTS

Detection of protease activity in BLV. To detect protease activity, the virus was incubated under the conditions used previously for the mouse protease assay (28). The optimum pH for protease activity was determined by incubating 30-μg portions of purified virus at various pHs in TD buffer containing 0.5% Nonidet P-40 for 16 h at 22°C in the presence or absence of the exogenous substrate Gz-MSV Pr65^{gag}. It appeared reasonable to use Gz-MSV Pr65^{gag} as the BLV protease substrate because the COOH-terminal amino acid sequence of the BLV *gag* proteins (-Pro-Ala-Ile-Leu-COOH) has been shown to be similar to the sequences of the mouse *gag* proteins (13, 15). When purified BLV alone was incubated under these conditions, no precursor-product relationship like that seen in the mouse system was observed (29, 30). However, when Gz-MSV Pr65^{gag} was added to BLV preparations in the assay described above, the Pr65^{gag} content decreased and there was a concomittant increase in the smaller proteins (data not shown). The activity was highest around pH 7.0 to 7.2. Thus, we decided to use Gz-MSV Pr65^{gag} as a substrate in BLV protease assays to monitor activity in the procedures developed for the purification of the enzyme.

TABLE 1. Amino acid composition of BLV protease[a]

Amino acid	No. of residues per protein determined	Predicted no. of residues[b]
Asp	10.32 (10)	10
Thr	4.86 (5)	5
Ser	7.75 (8)	9
Glu	13.29 (13)	11
Pro	12.07 (12)	14
Gly	10.03 (10)	9
Ala	9.56 (10)	9
Val	10.02 (10)	11
Met	0.66 (1)	2
Ile	6.78 (7)	8
Leu	15.83 (16)	17
Tyr	2.79 (3)	3
Phe	2.08 (2)	2
His	0.95 (1)	0
Lys	4.21 (4)	3
Arg	8.00 (8)	9
Cys	ND[c]	0
Trp	ND	4

[a] The determined molecular weight was 13,763, and the predicted molecular weight was 14,035.

[b] Amino acid composition predicted from the DNA sequence (18). Based on the determined NH_2 and COOH termini, the sequence from nucleotide 1677 to nucleotide 2054 was used for this comparison.

[c] ND, Not determined.

TABLE 2. N-terminal amino acid analysis of BLV protease

Cycle	Amino acid	Amt recovered (pmol)[a]
1	Leu	210
2	Ser	26
3	Ile	203
4	Pro	155
5	Leu	167
6	Ala	146
7	Arg	65
8	X	
9	Arg	93
10	Pro	100
11	Ser	9
12	Val[b]	
13	Ala	90
14	Val	
15	Tyr	79
16	Leu	94
17	Ser	5
18	Gly	37
19	Pro	56
20	Trp	11
21	Leu	83
22	Gln	26
23	Pro	55
24	(Ser)	2
25	Gln	55
26[c]		
27	Gln	43
28	Ala	21
29	Leu	37
30	Met	12
31	Leu	46

[a] Based on the phenylthiodhydantoin amino acid derivatives.

[b] Valine assignments were qualitive only due to the presence of an artifactual peak in the chromatographic analysis of the phenylthiodhydantoin amino acids.

[c] Sample was lost during workup.

Purification of the protease. The protease (molecular weight, 14,000 [14K]) eluted from the Sephacryl-200 column (see Materials and Methods) was further purified by RP-HPLC, using a shallow acetonitrile gradient on a μBondapak C_{18} column (Fig. 1). Fractions (5 ml) were collected and lyophilized for determinations of purity and activity. The peak activity was eluted in fraction 11 at an acetonitrile concentration of about 52 to 54%. This activity was clearly separated from proteins p12, p10, and p15. The purified protein produced a single band in SDS-PAGE gels (Fig. 1, fraction 11). The amount of total protein recovered in RP-HPLC fraction 11 was 35 μg. The total recovery of enzyme activity was 40%, as determined from a 50% reduction in the Pr65^{gag} band density after 16 h of incubation (28). Relatively large amounts of activity were lost during extraction from the acetone powder of the virus. This was probably caused by the hydrophobic nature of the protein, as indicated by the position of its elution in the acetonitrile gradient. However, this step was essential to remove the bulk of hydrophobic and low-molecular weight contaminants. To identify the protein responsible for the protease activity, the purified protease fraction from RP-HPLC was incubated with Gz-MSV Pr65^{gag}. As shown in Fig. 2, the activity was found in the 14K protein (fraction 11). In this fraction Pr65^{gag}

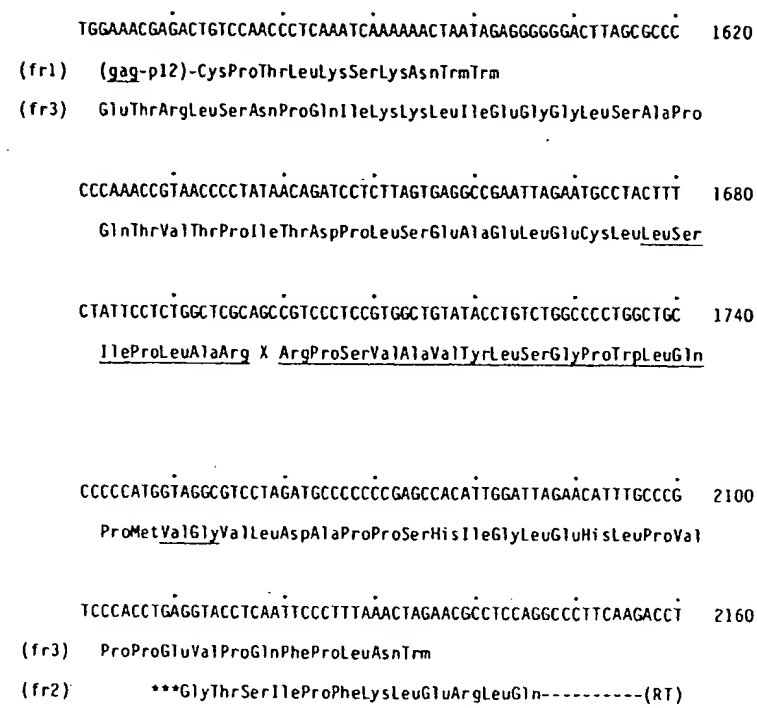


FIG. 3. Alignment of the determined NH_2- and COOH-terminal amino acid sequences (underlined) of the protease with the amino acid sequence deduced from the DNA sequence (18). RT, Reverse transcriptase; fr1, frame 1.

was cleaved into 44K, 40K, and 20K major bands (all doublets), including minor proteins after incubation (Fig. 2, lane 5). The BLV protease purified in this way apparently did not process Gz-MSV $Pr65^{gag}$ to the usual final products (p15, p30, p12, and p10). It remains to be determined which sites in Gz-MSV $Pr65^{gag}$ are cleaved by the BLV protease. This could be accomplished by an NH_2- and COOH-terminal analysis of each product.

Chemical analysis of the protease. The amino acid composition data for the protease are shown in Table 1. The total number of amino acids, not including cysteine and tryptophane, was 126, and the molecular weight calculated from these data was 13,763. The composition which we determined is in good agreement with the composition predicted from the nucleotide sequence (Table 1). To determine the NH_2-terminal amino acid sequence of the purified protease recovered from fraction 11, 0.25 nmol of protein was subjected to analysis on a gas phase sequenator. The amino acids identified at each cycle (of the first 31 cycles) are shown in Table 2, together with the quantitative yields. To determine whether the protease is virus encoded, we aligned the NH_2-terminal amino acid sequence with the amino acid sequence deduced from the nucleotide sequence of the DNA. As shown in Fig. 3, the protease amino acid sequence starts with a leucine residue encoded by the triplet at nucleotides 1677 to 1679 in the *pol* gene (18), 70 nucleotides downstream from the end of *gag* gene, and located, as expected, in a reading frame different from that of *gag* (18). The following two lines of evidence suggested that the protease COOH terminus is at nucleotides 2052 to 2054 and not at nucleotides 2130 to 2132 of this open reading frame adjacent to the termination codon (UAG): (i) the molecular weight estimated from the SDS-PAGE profile and the amino acid composition data (Fig. 1 and Table 1) was about 14,000; and (ii) the data from a preliminary analysis showed that the COOH-terminal sequence of this protein was -Val-Gly-COOH.The protein and nucleotide sequence data together clearly showed that the protease-encoding region located in the genome from nucleotide 1677 to nucleotide 2056 is in a different frame than *gag* and the putative reverse transcriptase gene.

In vitro processing of BLV *gag* precursors by the protease. To determine whether the BLV protease cleaves the BLV polyprotein precursor accurately, we used the polyprotein substrates synthesized in vitro. BLV RNA-directed cell-free protein synthesis yielded three major proteins, designated the 66K, 44K, and 14K proteins. The 66K and 44K proteins were related to the *gag* gene, as demonstrated by immunoprecipitation with anti-p24 antiserum (Fig. 4B, lane 5, and Fig. 4C, lane 7). The 14K protein is not considered here; it is described elsewhere (31). After the immune complexes (antigen-antibody-protein A-Sepharose complexes) of the 66K and 44K proteins were incubated with the partially purified protease under the assay conditions described above, the cleavage products were analyzed by SDS-PAGE. As shown in Fig. 4, the same sizes of BLV *gag* proteins (p24, p15, p12, and p14 [protease]) were generated by the protease treatment (see below) (Fig. 5). Additional bands, labeled bands m and n, which were also present in the purified virus preparation, were detected as cleavage products (Fig. 4C, lane 9). We identified band m as p10, an NH_2-terminal polypeptide of *gag* polyproteins. From these results, we concluded that the purified protease processes the BLV *gag* precursors in a specific manner.

DISCUSSION

BLV protease was purified, and its NH_2- and COOH-terminal amino acid sequences were determined. Our results show that the gene for the BLV protease (M_r, 14,000) is located between the *gag* gene and the putative reverse

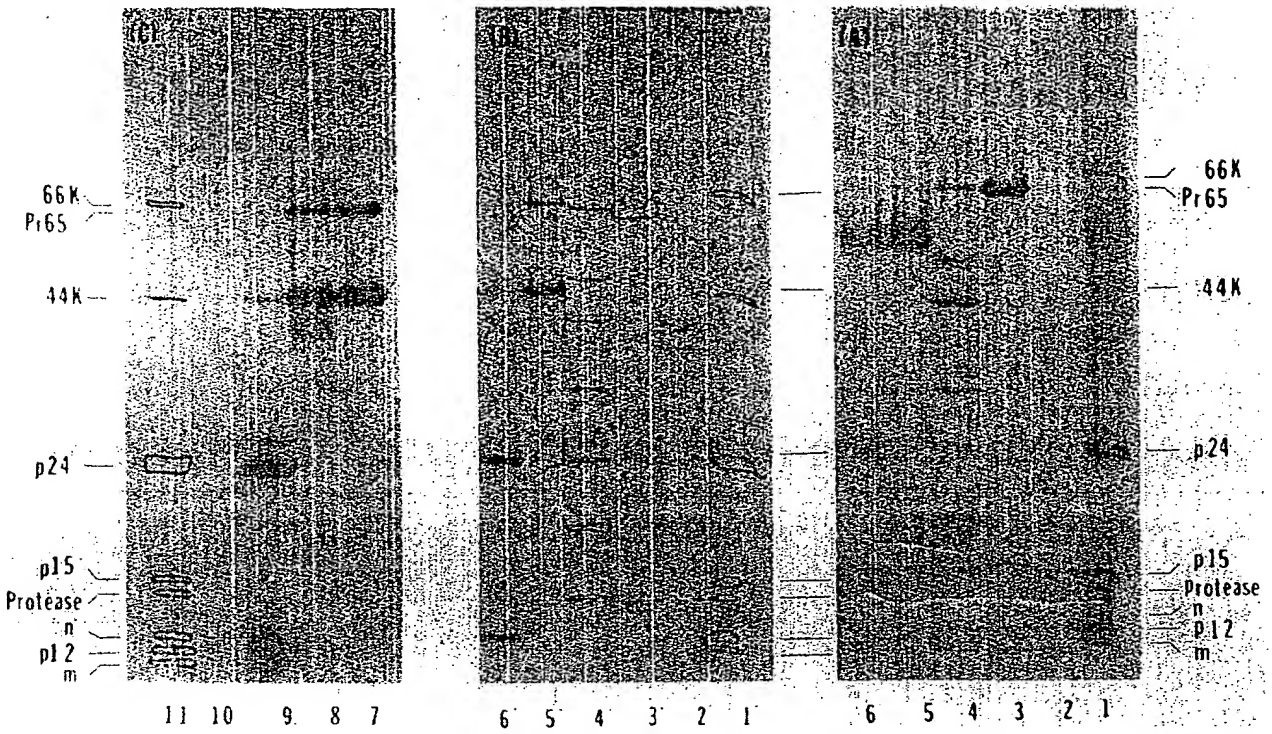


FIG. 4. In vitro cleavage of the BLV *gag* precursor proteins with the BLV protease. Whole incubation mixtures (100 μl; labeled with L-[^{35}S]methionine or L-[^{35}S]cysteine) of the BLV RNA-directed in vitro translation system were immunoprecipitated with anti-p24 rabbit antiserum (see text). The immune complex was then incubated with the BLV protease fraction under the assay conditions described in the text. After incubation, proteins were analyzed by SDS-PAGE and stained with Coomassie brilliant blue R-250, and then autoradiography was performed by using sodium salicylate. (A) Stained gel of the L-[^{35}S]methionine labeling experiment. Lane 1, BLV whole virus (40 μg); lane 2, partially purified protease (the RP-HPLC fraction used in this experiment contained some p24 in addition to protease p14); lane 3, Gz-MSV (7.5 μg); lane 4, Gz-MSV plus partially purified protease; lane 5, the immune complex alone; lane 6, the immune complex plus partially purified protease. (B) Autoradiogram of the L-[^{35}S]methionine labeling experiment. The stained gel (panel A) was exposed to X-ray film. Lanes 1' to 4', Tracings of lanes 1 to 4 in the stained gel (panel A); lane 5', the immune complex alone; lane 6', the immune complex plus partially purified protease. (C) Autoradiogram of the L-[^{35}S]cysteine labeling experiment. Lane 7, The immune complex alone (no incubation); lane 8, the immune complex alone (incubated); lane 9, the immune complex plus partially purified protease; lane 10, partially purified protease; lane 11, BLV whole virus (40 μg) (tracings of the stained gel).

transcriptase gene and is coded in a reading frame that is different from the reading frames of the *gag* and reverse transcriptase genes (18). In both Moloney murine leukemia virus and feline leukemia virus (FeLV) the protease gene is located in the *pol* gene in the same reading frame as the *gag* gene (27, 28); and in the avian retroviral system, the protease gene is located in the 3' region of the *gag* gene (22). Thus, in BLV the organization of the protease gene in the viral genome is different from the organization in both murine leukemia virus and FeLV systems and avian systems. However, within the protease-encoding region the translated BLV sequence shows homology with the avian and mammalian type C virus sequences over the entire length of the region. The BLV protease is most closely related to mouse protease (18).

Based on an NH_2- and COOH-terminal analysis of the *gag* proteins, we suggested previously (13, 27) that the cleavage sites in human T-cell leukemia virus and FeLV *gag* precursors are very similar to the cleavage sites in the mouse *gag* polyprotein. The cleavage sites in BLV are also similar in the COOH-terminal amino acid sequences of each *gag* protein to the cleavage sites in the above-mentioned retroviruses (4, 13, 14, 27). Based on this similarity, we used Gz-MSV $Pr65^{gag}$ as the substrate for the BLV protease during purification steps. Purified BLV protease was capable of cleaving the mouse *gag* precursor but apparently did not readily produce the correct final products (Fig. 2 and 4). However, we found that purified or partially purified BLV protease cleaved in vitro synthesized BLV polyproteins into the expected viral *gag* components without yielding other major products. Furthermore, when we used other proteins, such as bovine serum albumin or the mouse immunoglobulin heavy chain, as the protease substrate, no cleavage was observed. The exact specificity of the in vitro protease cleavage remains to be determined.

Another interesting question is the mechanism of the synthesis of BLV protease. In murine leukemia virus and FeLV, the protease is synthesized by readthrough of an amber termination codon at the end of the *gag* gene to produce a *gag-pol* fusion polyprotein. After being synthesized under translational control in a quantity that is small relative to the amount of *gag* precursor ($Pr65^{gag}$), $Pr180^{gag\text{-}pol}$ is processed by proteolytic cleavage, and the protease is generated as Thr-Leu-Asp-Asp (from the *gag* gene)-Gln-(*pol* gene product). An intriguing hypothesis has been proposed for BLV protease synthesis (18); according to this hypothesis a −1 frameshift might occur near the end of the *gag* gene. In that region, a guanine-cytosine-rich hairpin structure and six consecutive adenine residues are found; both characteristic sequences have been thought to be

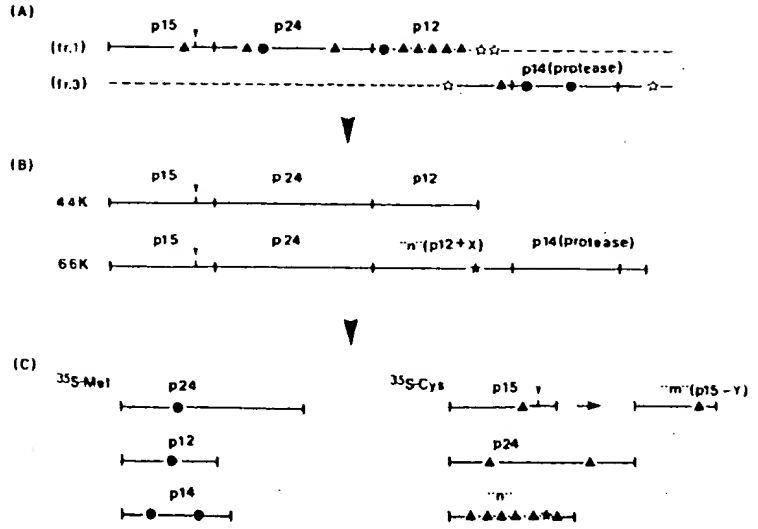


FIG. 5. Model for the synthesis of *gag* and protease proteins. Based on the DNA sequence data and the results obtained from the cleavage patterns in the processing of the *gag*-related 66K and 44K polyprotein precursor by the protease, a model for the origin of the BLV protease was made. (A) Coding frame of the *gag* and protease in the BLV genome. Symbols: ☆, termination codon; ●, methionine residue found in the DNA sequence; ▲, cysteine residue found in the DNA sequence. fr., Frame. (B) Translated product from coding frames 1 and 3. The partitions determine the NH_2 and COOH termini (cleavage sites) of each protein. Symbols: ★, point where a frameshift (from frame 1 to frame 3) could occur; ▼, another possible cleavage by BLV protease. (C) Final cleavage products, detected by labeling with either L-[^{35}S]methionine or L-[^{35}S]cysteine. All of these products were apparently identical to the products found in the purified virus. Both p12 (six cysteines) and band "n" (p12 + x; seven cysteines) are seen in Fig. 4. Only "n" is indicated above.

important for translational control of mRNA in procaryotes as well as eucaryotes (1, 5, 10). Not only the precise localization and reading frame assignment as described in this paper, but also the results of our biochemical studies with in vitro synthesized proteins are also compatible with this hypothesis. As reported previously by Ghysdale et al. (7) and as shown in Fig. 4, the 66K and 44K proteins were *gag*-related precursors. If frameshift suppression of the *gag* termination codon is the mechanism involved in the synthesis of the 66K polyprotein, we can expect the final cleavage products which we show in Fig. 5C. These products were obtained after incubation of the in vitro synthesized 66K and 44K polyproteins labeled with either [^{35}S]methionine or [^{35}S]cysteine. As shown in Fig. 4, we detected p14 protease, p24, and p12 labeled with [^{35}S]methionine (Fig. 4B, lane 6') and p24, p15, band n, p 12, and band m labeled with [^{35}S]cysteine. p24, p15 and p12 are known structural proteins of BLV. Band n, which migrated slower than p12 (Fig. 4), is what would be expected to be a product of [^{35}S]cysteine-labeled 66K protein after incubation with protease consisting of a major portion of p12 plus downstream sequences in the protease reading frame. As for the fastest-migrating band (band m) detected in the [^{35}S]cysteine labeling experiment, we determined that this protein was p10, an NH_2-terminal piece of p15. This protein was also found to be myristylated in the virus (unpublished data). Therefore, all of these products, including the band n and m proteins, which were obtained after cleavage of the 66K and 44K precursors with the protease (Fig. 4), match our model (Fig. 5) for the synthesis and cleavage of *gag* and the protease. From these results it seems likely that the protease is synthesized by frameshift suppression of the *gag* termination codons.

The preliminary results of our in vitro translation experiments with lysine-specific tRNA ($tRNA_5^{Lys}$; a gift from B. J. Ortwerth, University of Missouri, Columbia) may support this conclusion. In these experiments, the synthesis of the *gag* polyprotein precursor 66K was promoted by the addition of lysine-specific tRNA to the BLV genomic RNA-directed cell-free translation system. The degree of promotion was proportional to the amount of the tRNA added, and about two- to threefold stimulation was obtained with 0.02 U of optical density at 260 nm. Synthesis of the 44K and 14K proteins was not promoted at this concentration of tRNA.

A direct amino acid sequence analysis of *gag*-protease fusion protein could elucidate further details of the translational control and processing of BLV proteins. The other possibility, that a spliced mRNA is required for the synthesis of the *gag*-protease polyprotein, cannot be ruled out. In any case, the diversity of translational control mechanisms at the *gag-pol* junction of retroviruses, such as avian virus, murine leukemia virus, FeLV, BLV, and human T-cell leukemia virus, provides useful models for examining features of gene expression in eucaryotic cells. It also remains to be determined whether the protease itself cleaves all of the sites in *gag*-protease (66K) polyprotein or whether a cellular protease triggers the initial cleavage event. As shown in Fig. 4, we incubated the 66K and 44K polyproteins after immunoprecipitation with antibody without added protease and did not observe lower-molecular-weight products. Despite our inability to demonstrate self-cleavage in vitro of the 66K protein which contains the protease sequence, the possibility of autocatalytic cleavage under natural conditions cannot be ruled out.

ACKNOWLEDGMENTS

We thank Young Kim and C. V. Hixson for excellent technical assistance, Jeannie Clarke and Sally Lockhart for preparation of the manuscript, and B. J. Ortwerth for the generous gift of lysine-specific tRNA.

This research was sponsored under Public Health Service contract N01-CO-23909 between the National Cancer Institute and Litton Bionetics, Inc.

LITERATURE CITED

1. Atkins, J., R. F. Gesteland, B. R. Rein, and C. W. Anderson. 1979. Normal tRNA promotes ribosomal frame shifting. Cell 18:1119–1131.
2. Burny, A., C. Bruck, H. Chantrenne, Y. Cleuter, D. Dekegel, J. Ghysdael, R. Kettmann, M. Leclercq, J. Leunen, M. Mammerickx, and D. Portetelle. 1980. Bovine leukemia virus: molecular biology and epidemiology, p. 231–289. *In* G. Klein (ed.), Viral oncology. Raven Press, New York.
3. Chamberlain, J. P. 1979. Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble flour, sodium salicylate. Anal. Biochem. 98:132–135.
4. Copeland, T. D., M. A. Morgan, and S. Oroszlan. 1983. Complete amino acid sequence of the nucleic acid binding protein of bovine leukemia virus. FEBS Lett. 156:37–40.
5. Fox, T. D., and B. Weiss-Brummer. 1980. Leaky +1 and −1 frameshift mutations at the same site in a yeast mitochondrial gene. Nature (London) 288:60–63.
6. Gazdar, A. F., L. A. Phillips, P. S. Sarma, P. T. Peebles, and H. C. Chopra. 1971. Presence of the sarcoma genome in a "non-infectious" mammalian virus. Nature (London) New Biol. 234:69–72.
7. Ghysdale, J., R. Kettmann, and A. Burny. 1979. Translation of BLV virion RNAs in heterologous protein-synthesizing systems. J. Virol. 29:1087–1098.
8. Henderson, L. E., T. D. Copeland, and S. Oroszlan. 1980. Separation of all amino acid phenylthiohydantoins by high-performance chromatography on phenylalkyl support. Anal. Biochem. 102:1–7.
9. Hewick, R. M., M. M. Hunkapiller, L. E. Hood, and W. J. Dryer. 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. 254:7990–7997.
10. Kohno, T., and J. R. Roth. 1978. A salmonella frameshift suppression that acts at runs of A residues in the messenger RNA. J. Mol. Biol. 126:37–52.
11. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
12. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. Oroszlan, S., and T. D. Copeland. 1985. Primary structure and processing of *gag* and *env* gene products of human T-cell leukemia viruses HTLV-I_{CR} and HTLV-I_{ATK}. Curr. Top. Microbiol. Immunol. 115:221–233.
14. Oroszlan, S., T. D. Copeland, L. E. Henderson, J. R. Stephenson, and R. V. Gilden. 1979. Amino terminal sequence of bovine leukemia virus major internal protein: homology with mammalian type C virus p30s. Proc. Natl. Acad. Sci. USA 76:2996–3000.
15. Oroszlan, S., L. E. Henderson, J. R. Stephenson, T. D. Copeland, C. W. Long, J. N. Ihle, and R. V. Gilden. 1978. Amino- and carboxyl-terminal amino acid sequences of proteins coded by *gag* gene of murine leukemia virus. Proc. Natl. Acad. Sci. USA 75:1404–1408.
16. Oroszlan, S., M. G. Sarngadharan, T. D. Copeland, V. S. Kalyanaraman, R. V. Gilden, and R. C. Gallo. 1982. Primary structure analysis of the major internal protein p24 of human type C T-cell leukemia virus. Proc. Natl. Acad. Sci. USA 79:1291–1294.
17. Poiesz, B. J., F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna, and R. C. Gallo. 1980. Isolation of type C retrovirus particles from cultured and fresh lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. USA 77:7415–7419.
18. Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden. 1985. The *gag* and *pol* genes of bovine leukemia virus: nucleotide sequence and analysis. Virology 142:357–377.
19. Rice, N. R., R. M. Stephens, D. Couez, J. Deschamps, R. Kettmann, A. Burny, and R. V. Gilden. 1984. The nucletide seuquence of the *env* gene and post-*env* region of bovine leukemia virus. Virology 138:82–93.
20. Sagata, N., T. Yasunaga, K. Tuzuku-Kawamura, Y. Ohishi, Y. Ogawa, and Y. Ikawa. 1985. Complete nucleotide sequence of the genome bovine leukemia virus: its evolutionary relationship to the other retroviruses. Proc. Natl. Acad. Sci. USA 82:677–681.
21. Schultz, A. M., T. D. Copeland, and S. Oroszlan. 1984. The envelope proteins of bovine leukemia virus: purification and sequence analysis. Virology 135:417–427.
22. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869.
23. Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida. 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome in integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA 80:3618–3622.
24. Van der Maaten, M. J., and J. M. Miller. 1976. Serological evidence of transmission of bovine leukemia virus to chimpanzees. Vet. Microbiol. 1:351–357.
25. Wray, W., T. Boulikas, V. P. Wray, and R. Hancock. 1981. Silver staining of proteins in polyacrylamide gels. Anal. Biochem. 118:197–203.
26. Yoshida, M., I. Miyoshi, and Y. Hinuma. 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl. Acad. Sci. USA 79:2031–2035.
27. Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan. 1985. Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. J. Virol. 55:870–873.
28. Yoshinaka, Y., K. Katoh, T. D. Copeland, and S. Oroszlan. 1985. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA 82:1618–1622.
29. Yoshinaka, Y., and R. B. Luftig. 1977. Properties of a P70 proteolytic factor of murine leukemia viruses. Cell 12:709–719.
30. Yoshinaka, Y., and R. B. Luftig. 1982. P65 of Gazdar murine sarcoma viruses contains antigenic determinants from all four of the murine leukemia virus (MuLV) *gag* polypeptides (p15, p12, p30, and p10) and can be cleaved *in vitro* by the MuLV proteolytic activity. Virology 118:380–388.
31. Yoshinaka, Y., and S. Oroszlan. 1985. Bovine leukemia virus post-envelope gene coded protein: evidence for expression in natural infection. Biochem. Biophys. Res. Commun. 131:347–354.

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 807–811, February 1989
Biochemistry

Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides

(polyprotein processing/peptide cleavage/acquired immunodeficiency syndrome)

HANS-GEORG KRÄUSSLICH†, RICHARD H. INGRAHAM‡, MARK T. SKOOG‡, ECKARD WIMMER†, PETER V. PALLAI‡, AND CAROL A. CARTER†

†Department of Microbiology, State University of New York at Stony Brook, Stony Brook, NY 11794; and ‡Boehringer Ingelheim Pharmaceuticals Inc., 90 East Ridge, P.O. Box 368, Ridgefield, CT 06877

Communicated by Howard L. Bachrach, October 19, 1988 (received for review August 19, 1988)

ABSTRACT **Retroviral capsid proteins and replication enzymes are synthesized as polyproteins that are proteolytically processed to the mature products by a virus-encoded proteinase. We have purified the proteinase of human immunodeficiency virus (HIV), expressed in *Escherichia coli*, to ≈90% purity. The purified enzyme at a concentration of ≈20 nM gave rapid, efficient, and specific cleavage of an *in vitro* synthesized *gag* precursor protein. Purified HIV proteinase also induced specific cleavage of five decapeptide substrates whose amino acid sequences corresponded to cleavage sites in the HIV polyprotein but not of a peptide corresponding to a cleavage site in another retrovirus. Competition experiments with different peptides allowed a ranking of cleavage sites. Inhibition studies indicated that the HIV proteinase was inhibited by pepstatin A with an IC_{50} of 0.7 μM.**

The capsid and nonstructural proteins of all retroviruses, including human immunodeficiency virus (HIV), are synthesized as polyprotein precursors that are proteolytically processed to the mature viral proteins by a virus-encoded, virus-associated proteinase (for a review see ref. 1). Viral proteinases (PR; for the new nomenclature for common retroviral proteins see ref. 2) have been purified from virions and biochemically characterized for a number of avian (3, 4) and mammalian (5–7) retroviruses. These enzymes share limited amino acid sequence homology with members of the aspartic proteinases (8, 9) and invariably contain the sequence Asp-Thr(Ser)-Gly corresponding to the catalytic center of cellular aspartic proteinases. However, retroviral enzymes are much smaller than cellular aspartic proteinases and contain only a single homologous catalytic center.

The proteolytic activity of HIV has been mapped to a 11-kDa protein that is encoded immediately upstream of the viral reverse transcriptase (RT) and appears to be generated by autocatalytic release from a larger precursor (10–12). Replication of infectious HIV particles is entirely dependent on the generation of active PR, and a mutation of the putative active site Asp residue in the PR gene resulted in the production of "immature," noninfectious particles consisting of uncleaved precursor proteins (13). Although genetic and biochemical evidence has mapped the proteolytic activity to this specific segment of the HIV genome, the mechanisms of PR formation and action are still unknown. Is dimerization of the enzyme a prerequisite for function (14) and, if so, how are the initial cleavages performed? Does the enzyme require activation or relief of inactivation or is processing confined to a specific (virus-induced?) compartment (for a review see ref. 1)? In addition to providing insight into the biology of retrovirus replication, studies of the enzymatic functions of HIV have recently received much attention as these enzymes are potential targets for antiviral drugs. HIV PR, like the proteinases of other human pathogenic viruses, such as poliovirus, appears sufficiently distinct from cellular endopeptidases that inhibitors of the viral enzyme may not be harmful to the host. These studies require sufficient quantities of pure, active enzyme and functional assays in a peptide-based system as well as on a natural substrate.

We report here purification of enzymatically active HIV PR from an *Escherichia coli* expression system. The purified enzyme can cleave its natural substrate, obtained by translation *in vitro*, as well as five different decapeptides designed according to cleavage sites in the HIV polyprotein. Processing of these peptides occurred specifically at the site that is also cleaved *in vivo*, and a non-HIV-specific peptide was not cleaved by HIV PR. Competition experiments with different peptide substrates allowed ranking of the peptides according to the relative rates of cleavage.

MATERIALS AND METHODS

Purification of HIV PR. Construction of plasmid pHIVproPII and expression in *E. coli* BL21(DE3) has been described (12). In the present study we used BL21(DE3) containing the T7 lysozyme gene on a chloramphenicol-resistance plasmid [BL21(DE3)plysS] as the expression strain. T7 lysozyme has been shown to inhibit T7 RNA polymerase (15), thus providing a more stringent control of base line expression. Bacteria transformed with pHIVproPII were grown in M9CA medium (16) with ampicillin and chloramphenicol in a 14-liter fermentor (Microferm MMF-14, New Brunswick Scientific, New Brunswick, NJ), essentially as described (17). Bacteria were collected by centrifugation and stored as a wet paste at −80°C.

For each purification, bacterial paste was thawed in 50 mM 2-(*N*-morpholino)ethanesulfonic acid (Mes), pH 6.5/0.1 M NaCl/10 mM $MgCl_2$/1 mM EDTA and lysed in a French pressure cell at 75 MPa. The lysate was centrifuged for 15 min at 10,000 × *g* and 60 min at 200,000 × *g*, the supernatant was collected and made 10 mM in EDTA, and $(NH_4)_2SO_4$ was added to 50% saturation. The precipitate was redissolved in 50 mM Tris·HCl, pH 8.0/30 mM NaCl/1 mM EDTA and layered on a DEAE-cellulose column (Whatman DE52) equilibrated with the same buffer. The flow-through was adjusted to pH 6.5, made 1 M in $(NH_4)_2SO_4$, and layered on a hexylagarose (Sigma) column equilibrated with 1 M $(NH_4)_2SO_4$ in 50 mM Mes, pH 6.5/1 mM EDTA (buffer A).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV, human immunodeficiency virus; TFA, trifluoroacetic acid; PR, protease; RT, reverse transcriptase.

The column was washed with the same buffer and with 0.85 M $(NH_4)_2SO_4$ in buffer A and was eluted with 50 mM $(NH_4)_2SO_4$ in buffer A. Protein was precipitated with $(NH_4)_2SO_4$, redissolved in 50 mM Mes, pH 6.5/150 mM NaCl/1 mM EDTA (buffer B), and loaded on a 60 × 1.6 cm column of Sephadex G-50 fine (Pharmacia) in buffer B at 9 ml/hr. Peak fractions were precipitated with $(NH_4)_2SO_4$, redissolved in buffer B, and loaded on a Superose 12 HR10/30 column (Pharmacia) at 0.5 ml/min.

In a control experiment BL21(DE3)plysS were transformed with plasmid pHIVproP (12). Purification was performed as described above except that the final Superose 12 and the hexylagarose step were omitted. Instead, the DEAE flow-through was precipitated by addition of $(NH_4)_2SO_4$ and the precipitate was redissolved in buffer B and loaded on the Sephadex G-50 column.

***In Vitro* Transcription and Translation.** Synthetic RNAs were transcribed with T7 RNA polymerase *in vitro* (18) from plasmids pHIVg/pII and pHIV FSII (12). Both plasmids contain the HIV cDNA sequence from nucleotides 221 to 2129 but pHIV FSII has a 4-base-pair insertion at nucleotide 1640 of the HIV cDNA leading to a switch from the *gag* reading frame to the *pol* reading frame (12). Synthetic RNAs (final concentration, 50 μg/ml) were translated in rabbit reticulocyte lysate (Promega Biotec) for 60 min at 30°C as described by the supplier.

Synthesis and Purification of Peptides. Peptides were synthesized by the Merrifield method on an Advanced Chemtech automated synthesizer and were cleaved from the resin by liquid HF at 0°C in the presence of anisole and dimethyl sulfide as scavengers. Dried peptide resin mixtures were washed with diethylether and peptides were extracted with water and 80% acetonitrile in water. The extracts were lyophilized and the crude peptides were purified by reversed-phase HPLC using 0.1% aqueous trifluoroacetic acid (TFA)/acetonitrile-based mobile phases. The lyophilized products were characterized by fast atom bombardment mass spectrometry, HPLC, and amino acid analysis.

Peptide Cleavage by HIV PR. Reactions were performed at 30°C in 20 or 40 μl of 50 mM sodium phosphate, pH 6.0/25 mM NaCl/5 mM EDTA/1 mM dithiothreitol with 0.44 mM decapeptide as substrate and 2 μl of partially purified PR (in purification buffer B). The reaction was quenched with a 4-fold excess of 0.1% TFA and frozen. Reaction products were analyzed by reversed-phase HPLC using a Vydac C_{18} analytical column with a gradient from 95% A/5% B to 100% B (A is 0.1% TFA in water; B is 0.09% TFA in 60% acetonitrile/40% water) in 30 min at a flow rate of 1 ml/min. Absorbance was monitored at 215 nm. Cleavage products were identified by sequencing on an Applied Biosystems model 477A sequencer equipped with a model 120A phenylthiohydantoin analyzer. Fast atom bombardment mass spectrometry was carried out on a Kratos MS-80 RFAQ double-focusing mass spectrometer.

Competition Experiments. Reactions were carried out at 37°C in buffer consisting of 50 mM Mes, 25 mM NaCl, and 5% dimethyl sulfoxide at pH 6.0. When the substrate contained methionine or cysteine residues, 1 mM dithiothreitol was included. Substrate concentration was 250 μM BI-P-136 and 250 μM another decapeptide. Reactions were started by adding 5–10 μl of enzyme in purification buffer B. For each time point a 20-μl sample was removed and added to 0.1 ml of 2% TFA in water. Peptides were separated on a Nucleosil C_{18} column with a 26-min linear gradient at 1 ml/min from 9% to 56% acetonitrile in water with 0.05% TFA and elution was monitored at 210 nm. For each substrate and its products, the area under the combined peaks was independent of extent conversion of the substrate. Thus, substrate and products were detected and recovered with equal efficiencies and comparison of areas yielded extent conversion.

When two substrates compete for the active site of an enzyme, kinetic analysis yields the ratio of V_{max}/K_m (19) rather than comparisons of either V_{max} or K_m individually. The rate constant ratios were determined from the extent of conversion of each substrate using the equation $(V_{max}/K_m)_1/(V_{max}/K_m)_2 = \log(1 - F_1)/\log(1 - F_2)$, where F is the fraction of substrate that is converted to products (20).

RESULTS

Purification of Biosynthetic HIV PR. We (12) and others (10, 11, 13, 21, 22) have recently reported expression of active HIV PR in *E. coli*. In our initial studies (12) we demonstrated that HIV PR was produced in *E. coli* BL21(DE3) transformed with the expression plasmid pHIVproPII, which contains the 5′ terminal part of the *pol* reading frame of HIV under the control of upstream elements of bacteriophage T7 gene *10*. The mature enzyme (99 amino acids; refs. 10 and 22) was released by autocatalytic cleavage of the precursor (12). The expression level for HIV PR was low and the protein could not be identified on a Coomassie blue-stained gel of bacterial lysates (Fig. 1*a*, lane 1). Large quantities of bacterial cells were grown in a bench-top fermentor because of this low-expression level.

We followed purification of the enzyme by immunoblot analysis using a polyclonal antiserum against HIV PR. This antiserum detected a single protein in lysates of induced bacteria (Fig. 1*b*, lane 3) and did not react with lysates that did not carry the expression plasmid (Fig. 1*b*, lane 1). The majority of HIV PR was soluble after high-speed centrifugation (Fig. 1*a*, lanes 2 and 3). The enzyme was quantitatively precipitated by $(NH_4)_2SO_4$ at 50% saturation (Fig. 1*a*, lane 4) and passed through a DEAE-cellulose column at low ionic strength. After this step, the enzyme could be visualized on a stained gel (Fig. 1*a*, lane 5) and was determined by laser densitometry to be ≈1% of total protein. The material was

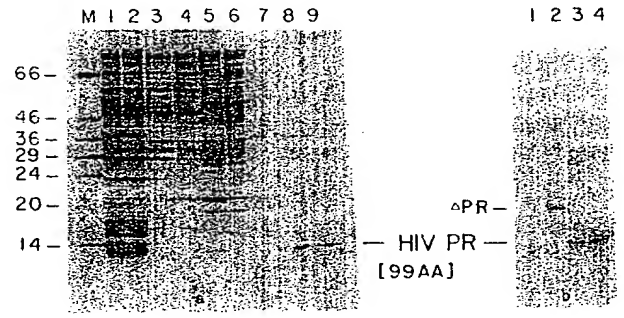


FIG. 1. Purification of HIV PR. (*a*) Aliquots taken at each step of the purification protocol were separated by NaDodSO$_4$/10–20% polyacrylamide gel electrophoresis and proteins were stained with Coomassie blue. AA, amino acids. Lane M refers to a marker lane containing 1 μg each of seven molecular size standards (indicated in kDa). In lanes 1–6, 20 μg of protein (determined using the Bio-Rad assay) was loaded, whereas lanes 7–9 contained ≈0.5 μg, ≈5 μg, and ≈1.5 μg, respectively. Lanes: 1, lysate of induced bacteria carrying pHIVproPII; 2, S-10 supernatant; 3, S-200 supernatant; 4, 50% $(NH_4)_2SO_4$ precipitate; 5, DEAE flow-through; 6, hexylagarose eluate; 7, pooled peak fractions from G-50 column; 8, $(NH_4)_2SO_4$ precipitate of G-50 pool; 9, pooled peak fractions from Superose 12 column. (*b*) Immunoblot analysis of bacterial fractions. Twenty micrograms of lysate from bacteria carrying the vector plasmid (lane 1), the deletion plasmid pHIVproP (lane 2), or the expression plasmid pHIVproPII (lane 3) and 1.5 μg of purified HIV PR (lane 4) were analyzed by gel electrophoresis and proteins were transferred to nitrocellulose for 2 hr at 0.2 A. The nitrocellulose paper was then probed with a polyclonal antiserum against HIV PR. Detection was with goat anti-IgG (rabbit), coupled to alkaline phosphatase (Tago), and with indolyl phosphate/nitro blue tetrazolium as the substrate/indicator system, essentially as described (23).

then passed over a hexylagarose column (Fig. 1*a*, lane 6) followed by molecular size chromatography on a Sepharose G-50 column (Fig. 1*a*, lanes 7 and 8) and a Superose 12 column. After this multistep purification, HIV PR was ≈90% pure as determined by laser densitometry of the gel shown in Fig. 1*a*, lane 9. Since the initial expression was low it is difficult to assess the level of overall recovery of protein but we could isolate ≈25 μg of HIV PR from 10 g of wet *E. coli* cell paste. Purified HIV PR was stable at −80°C for several weeks and did not appreciably lose activity on repeated freeze-thaw cycles.

HIV PR obtained by this purification reacted strongly with the polyclonal antiserum in immunoblot analysis (Fig. 1*b*, lane 4). The N-terminal sequence of the protein was determined by automated Edman degradation on an Applied Biosystems model 477A protein sequencer, using an *o*-phthaldialdehyde cycle in the first step to block N-terminal residues present other than proline as described by the supplier. Sequencing of the first five residues yielded Pro-Gln-Ile-Thr-Leu, the sequence found at the amino terminus of HIV PR (10, 22).

Cleavage of HIV *gag* Precursor by Purified HIV PR. In our initial experiments we observed that crude extracts from bacteria carrying plasmid pHIVproPII induced cleavage of *in vitro* synthesized *gag* precursor pr53, whereas bacterial extracts carrying pHIVproP, which has a deletion of the 17 C-terminal codons of HIV PR, did not catalyze this reaction (12). We now used this trans assay to monitor activity of purified HIV PR.

Fig. 2, lane 4, shows the products of translation of a synthetic mRNA containing the coding region for HIV *gag* and, in a different reading frame, for PR. The primary translation product was the *gag* precursor pr53, which was completely stable to incubation in buffer alone (Fig. 2, lane 5; see also ref. 12). Incubation of the *gag* precursor with purified HIV PR (Fig. 2, lanes 7–13) gave rapid and efficient processing to HIV capsid proteins. In lane 7, ≈0.5 μM HIV PR was present in the incubation and in lane 8, ≈20 nM

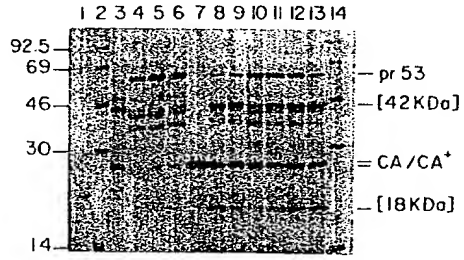


FIG. 2. Proteolytic processing of *in vitro* synthesized HIV *gag* precursor proteins. Samples of translation mixtures programed with no mRNA (lane 1), FSII RNA (lane 3), or g/pII RNA (lanes 4–13) either were mixed directly with sample buffer (lanes 1–3) or incubated with or without addition of bacterial fractions. Cleavage reactions were carried out in 20 μl final volume using 1.5 μl of translation mix as substrate and 1 μl of purified enzyme. Incubation was in 50 mM Mes, pH 6.0/20 mM NaCl/5 mM EDTA for 60 min at 30°C unless otherwise stated. Reactions were analyzed on 12.5% polyacrylamide/NaDodSO$_4$ gels. Kodak XAR-5 film was exposed to the dried gels for 16 hr. Lanes: 1, no mRNA; 2, mixture of ^{14}C-methylated proteins obtained from Amersham (sizes indicated in kDa); 3, FSII RNA; 4, g/pII RNA, no incubation; 5–13, g/pII RNA incubated with buffer alone for 1 hr (lane 5), partially purified extract from bacteria carrying plasmid pHIVproP for 1 hr (lane 6), ≈0.5 μM HIV PR for 1 hr (lane 7), or ≈20 nM HIV PR for 1 hr at 30°C (lane 8), 1 hr at 37°C (lane 9), 2 min at 30°C (lane 10), 5 min at 30°C (lane 11), 15 min at 30°C (lane 12), and 30 min at 30°C (lane 13); 14, marker lane as in lane 2. Note that a protein of ≈42 kDa was seen even without incubation with PR (lanes 4–6). This protein is distinct from the cleavage intermediate and is derived from internal initiation at Met-142 of the *gag* reading frame.

purified HIV PR was used. The lower concentration of enzyme gave almost complete cleavage of pr53 to yield the major capsid protein CA/Ca$^+$ (CA$^+$ specifies the C-terminally extended capsid protein p25; ref. 24) and two intermediates of ≈42 kDa and ≈18 kDa. Both intermediates were completely processed when 0.5 μM HIV PR was used. The 18-kDa protein probably corresponds to the C-terminal intermediate of *gag*, which is further processed to the nucleocapsid protein NC and protein p9 (24). By using a different gel system, we could detect a product that migrates at ≈8–10 kDa (data not shown). The 42-kDa protein probably corresponds to an intermediate containing the matrix protein MA and CA/CA$^+$ (12). The 42-kDa intermediate and CA comigrated with the corresponding products from an *in vitro* synthesized precursor containing equimolar amounts of *gag* and PR, which yielded efficient autocatalytic processing (Fig. 2, lane 3). Incubation of pr53 with material purified from bacteria carrying the deletion plasmid pHIVproP (12) did not result in any cleavage (Fig. 2, lane 6), although 50-fold more protein was used than in the experiment shown in lane 7.

We followed processing of pr53 with purified HIV PR in a time-course experiment (Fig. 2, lanes 10–13). Cleavage occurred very rapidly and cleavage products were already observed after a 2-min incubation (Fig. 2, lane 10). The cleavage assay was optimally performed at 30°C (Fig. 2, lanes 8 and 9). The pH optimum for cleavage was between pH 5.5 and 6.5, with pH >7 strongly inhibiting the activity, and the optimal salt concentration was between 20 and 100 mM NaCl (data not shown).

Cleavage of Peptide Substrates by Purified HIV PR. While the purification protocol was being developed, we tested the activity of partially purified HIV PR on synthetic peptide substrates. These peptides were modeled according to cleavage sites in the HIV *gag-pol* precursor. Fig. 3 shows that HIV PR induced cleavage of a decapeptide containing the sequence at the PR/RT cleavage site in the HIV *gag-pol* precursor (peptide BI-P-127; see Table 1). Two products, distinct from the parent peptide, were resolved by reversed-phase HPLC (Fig. 3 *b* and *c*). Sequence analysis of the peptide products confirmed that they had the sequences C-T-L-N-F and P-I-S-P-I, respectively, and therefore resulted from cleavage at the Phe/Pro bond. No other peptide material could be found in any other fraction. Integration of

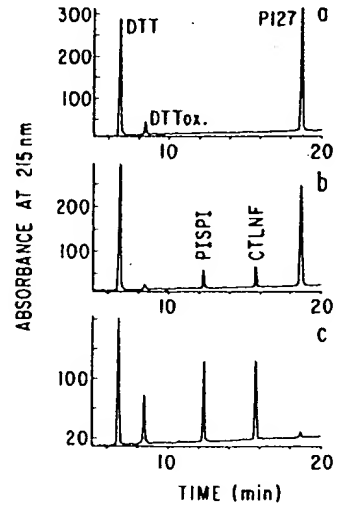


FIG. 3. Reversed-phase HPLC analysis of peptide BI-P-127 and HIV PR-generated cleavage products. The peptide was incubated for 4 hr in buffer alone (*a*) or for 1 hr (*b*) or 4 hr (*c*) with purified HIV PR. Monitoring was performed at 215 nm. DTT and DTTox. correspond to dithiothreitol and oxidized dithiothreitol, respectively. Cleavage products are identified by their amino acid sequence.

Table 1. Relative cleavage of HIV peptide substrates

Cleavage site*	Sequence					↓							Code	$(V_{max}/K_m)_{rel.}$†
	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'			
p6*/PR	V	S	F	N	F	*	P	Q	I	T	L	– NH_2	BI-P-136	1.00
CA^+/NC	T	A	T	I	M	*	M	Q	R	G	N	– NH_2	BI-P-140	0.20
MA/CA	V	S	Q	N	Y	*	P	I	V	Q	N	– NH_2	BI-P-138	0.07
CA/CA^+	K	A	R	V	L	*	A	E	A	M	S	– NH_2	BI-P-144	0.04
PR/RT	C	T	L	N	F	*	P	I	S	P	I	– NH_2	BI-P-127	0.03
RT/IN (avian)	Ac–T	F	Q	A	Y	*	P	L	R	E	A	– NH_2	BI-P-102	<0.005

IN, integrase protein.

*Cleavage sites within the HIV *gag-pol* polyprotein are designated according to the new nomenclature (2), except for the N-terminal product from the *pol* reading frame (p6*), for which there is no new name. CA^+ specifies the C-terminally extended capsid protein p25 (24).

†Relative values of V_{max}/K_m were determined by using competition experiments. Each value is an average of at least three determinations and is reproducible to ±20%.

the absorbance peaks corresponding to the parent peptide and the two cleavage products showed complete conversion of parent to product and no losses were observed. To further support the specificity of this cleavage, peptide BI-P-127 was incubated with material purified from bacteria carrying the deletion plasmid (pHIVproP), which does not contain active HIV PR. This incubation did not result in any processing of the peptide, although 50-fold more protein was used than in the experiments with active HIV PR (data not shown).

In addition to peptide BI-P-127, four other decapeptides, corresponding to cleavage sites within the HIV polyprotein, were synthesized. Peptide BI-P-136, which contains the N-terminal cleavage site of HIV PR (Table 1), was specifically cleaved at the Phe/Pro bond to release fragments corresponding to the N terminus of the proteinase and the C terminus of the upstream *pol* sequences (p6*). Both cleavage products were identified by sequence analysis of materials in the respective peaks after HPLC analysis. Fast atom bombardment mass spectrometry showed total conversion of material in a peak of 1164 atomic mass units to material of 570 atomic mass units (C-terminal fragment) and of 613 atomic mass units (N-terminal fragment) corresponding to the parent peptide and the two products after cleavage at the Phe/Pro bond, respectively. Three more decapeptides corresponding to the N- and C-terminal cleavage sites of the capsid protein CA/CA^+ (24) were synthesized. These peptides contained a Tyr/Pro (BI-P-138), Leu/Ala (BI-P-144), or Met/Met (BI-P-140) dipeptide as scissile bond. Incubation of these peptides with HIV PR gave specific cleavage in all cases (Table 1), and all products were identified by sequence analysis.

In addition to these HIV-specific peptides we studied the activity of HIV PR on a peptide that did not contain any HIV cleavage site. Peptide BI-P-102 corresponds to the cleavage site between the α-subunit of RT and IN of avian sarcoma-leukosis virus (Table 1; ref. 25) and contains a Tyr/Pro dipeptide as scissile bond but was completely stable to incubation with HIV PR under conditions where complete cleavage of the HIV-specific peptides occurred (Table 1).

In a separate set of experiments, we determined the relative susceptibility of these peptides to cleavage by HIV PR. Relative values of V_{max}/K_m for all substrates (Table 1) were obtained in experiments where two peptides, BI-P-136 and another, were incubated simultaneously with HIV PR and the substrates competed for the active site. This method requires fewer experiments than determining absolute values for V_{max}/K_m for each substrate and is not affected by variations in enzyme activity in different experiments. Competition of peptide BI-P-136 with each of the other five peptides resulted in the relative values V_{max}/K_m shown in Table 1. The best substrate, BI-P-136, was 5-fold more active than any other peptide tested, and peptide BI-P-140, containing a Met/Met dipeptide as the scissile bond, was cleaved at a significantly faster rate than the other three HIV-specific peptides.

The approximate turnover number for HIV PR was estimated from the rate of substrate turnover and an estimate of the enzyme concentration. Cleavage of 250 μM BI-P-136 was complete within 20 min in the presence of ≈200 nM HIV PR. This corresponds to a turnover number of ≈1 s^{-1}, if we assume that the enzyme was saturated under these conditions and the observed rate thus represents V_{max}. This value is a lower limit for the turnover number since the enzyme could not possibly be saturated for the entire course of cleavage of BI-P-136.

Since HIV PR was shown to be inhibited by high concentrations of pepstatin A (12, 21, 26), we determined the effect of pepstatin A on peptide cleavage (using peptide BI-P-136 as substrate) by purified HIV PR. These experiments gave an IC_{50} value for pepstatin A of 0.7 μM.

DISCUSSION

In this communication we report purification of active HIV PR to 90% purity. The enzyme was expressed in bacteria to a relatively low level but it was largely soluble and could therefore be purified by making use of its basic net charge, hydrophobicity, and low molecular mass. These results are in contrast to a report showing partial purification of HIV PR from *E. coli* (21), in which the enzyme was highly insoluble. Moreover, these authors followed purification of the enzyme only by immunoblot analysis and the degree of purity of their partially purified proteinase was not stated.

The purified enzyme was shown to be active on an *in vitro* synthesized *gag* precursor and on decapeptide substrates. Processing of the structural precursor was very efficient, giving almost complete cleavage with ≈20 nM enzyme. Thus, HIV PR was significantly more active on its natural substrate than poliovirus 3C, for which a concentration of 25 μM was required to achieve equally good cleavage of its natural substrate (17). An efficient proteolytic enzyme is desirable for HIV since *gag* precursor and PR are produced in a ratio of about 10:1 (27) and every enzyme molecule must cleave at least 50 peptide bonds to achieve complete processing of the *gag* and *gag-pol* polyproteins.

Purified HIV PR induced cleavage of five decapeptides that contained sequences corresponding to cleavage sites within the HIV polyprotein. Specific cleavage of peptides was also observed in a very recent study using chemically synthesized HIV PR (26). However, no kinetic analysis was performed in this study and the experiments were carried out on peptides of variable length and at a different pH. These differences may account for the different "ranking" of peptides these authors observed. Moreover, the chemically synthesized PR seems to be considerably less active than PR purified from *E.*

coli. This may be due to the presence of a high percentage of inactive molecules that accumulated during chemical synthesis.

On the basis of amino acid comparisons between different viruses, a consensus sequence for retroviral cleavage sites has been proposed (28), suggesting a generally hydrophobic pattern of amino acids surrounding the cleavage site with Tyr (Phe)/Pro being the most frequently occurring P1 and P1' residues, respectively. Three of our five decapeptide substrates match this pattern of amino acids very closely, whereas the other two are dissimilar but still contain primarily hydrophobic residues surrounding the scissile bond. Additional substrate determinants are likely to reside in amino acids on both sides of the scissile bond but larger data sets are required to define these determinants. HIV PR did not cleave an oligopeptide containing a Tyr/Pro cleavage site that served as a substrate for PR of avian sarcoma-leukosis virus (25). On the other hand, avian PR could process several oligopeptides corresponding to HIV cleavage sites (25). It remains to be seen whether the substrate requirements for small peptides generally match those for cleavage of viral polyproteins. Accessibility of potential cleavage sites and structurally flexible contexts are obvious additional determinants.

Cleavage of different dipeptide bonds by one enzyme raises the possibility of differences in the relative susceptibility of these sites. This could be a regulatory feature in that processing may be required for activation or inactivation of functional domains within the polyprotein during and after assembly of the virion. In competition experiments we demonstrated that a peptide corresponding to the N-terminal cleavage site of PR was processed significantly faster than all other peptide substrates. These results suggest that this cleavage, which leads to separation of the structural proteins of the nucleocapsid from the replication enzymes, may be a very efficient first step in the processing pathway. A similar relative order of processing steps is observed in picornavirus replication where cleavage between the structural and nonstructural domains of the viral polyprotein is a very efficient first step in the processing cascade (1). Recently, it was shown that the CA/NC cleavage can occur at three different sites within a 16-amino acid sequence (24). Our results indicate that purified HIV PR can cleave at least two of these sites and that the Met/Met site may be the preferred cleavage site since processing of a peptide containing the Leu/Ala site was significantly slower.

Retroviral proteinases are believed to be aspartic proteinases and it is of interest in this regard that pepstatin A, which had been shown to inhibit polyprotein cleavage by retroviral enzymes at a very high concentration (>0.1 mM; refs. 9, 12, 21, and 26), inhibited peptide cleavage by purified HIV PR with an IC_{50} of 0.7 μM. This inhibitory effect is still very weak compared to a K_i for pepsin of 4.5×10^{-11} M (29) but is comparable to the effect of pepstatin A on renin (K_i = 0.1–1 μM; ref. 29), consistent with the classification of the retroviral enzymes as aspartic proteinases.

We thank J. J. Dunn and F. W. Studier for purified T7 RNA polymerase and BL21(DE3)plysS and P. Barr for HIV PR-specific antiserum. We are indebted to K. Harris, J. Sellmann, and H. Schneider for assistance in the purification and assays; F. Burkhardt and D. Freeman for peptide synthesis and purification; T. Seng for sequence analysis; P. Bax and G. Chow for assistance in the competition experiments; and K. McKellop for performing mass spectrometry. We also thank R. Cabelli and M. Reddy for helpful discussions and suggestions, A. Kameda for artwork, and C. Helmke for photography. This work was supported in part by Grant AI25993 from the National Institutes of Health. H.-G.K. is supported by Fellowship Kr906/1-3 from the Deutsche Forschungsgemeinschaft.

1. Kräusslich, H.-G. & Wimmer, E. (1988) *Annu. Rev. Biochem.* **57**, 701–754.
2. Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M. & Vogt, V. (1988) *J. Virol.* **62**, 1808–1809.
3. von der Helm, K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 911–915.
4. Dittmar, K. J. & Moelling, K. (1978) *J. Virol.* **28**, 106–118.
5. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1618–1622.
6. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) *J. Virol.* **55**, 870–873.
7. Yoshinaka, Y., Katoh, I., Copeland, T. D., Smythers, G. W. & Oroszlan, S. (1986) *J. Virol.* **57**, 826–832.
8. Toh, H., Ono, M., Saigo, K. & Miyata, T. (1985) *Nature (London)* **315**, 691–692.
9. Katoh, I., Yasunaga, Y., Ikawa, Y. & Yoshinaka, Y. (1987) *Nature (London)* **329**, 654–656.
10. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W. & Rosenberg, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8903–8906.
11. Graves, M. C., Lim, J. J., Heimer, E. P. & Kramer, R. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2449–2453.
12. Kräusslich, H.-G., Schneider, H., Zybarth, G., Carter, C. A. & Wimmer, E. (1988) *J. Virol.* **62**, 4393–4397.
13. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M. & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4686–4690.
14. Pearl, L. H. & Taylor, W. R. (1987) *Nature (London)* **329**, 351–354.
15. Moffatt, B. A. & Studier, F. W. (1987) *Cell* **49**, 221–227.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
17. Nicklin, M. J. H., Harris, K., Pallai, P. V. & Wimmer, E. (1988) *J. Virol.* 4586–4593.
18. van der Werf, S., Bradley, J., Wimmer, E., Studier, F. W. & Dunn, J. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2330–2334.
19. O'Leary, M. H. & Baughn, R. L. (1972) *J. Am. Chem. Soc.* **94**, 626–630.
20. Melander, L. & Saunders, W. H., Jr. (1987) *Reaction Rates of Isotopic Molecules* (Krieger, Melbourne, FL), p. 96.
21. Hansen, J., Billich, S., Schulze, T., Sukrow, S. & Moelling, K. (1988) *EMBO J.* **7**, 1785–1791.
22. Lillehoj, E. P., Salazar, F. H. R., Mervis, R. J., Raum, M. G., Chan, H. W., Ahmad, N. & Venkatesan, S. (1988) *J. Virol.* **62**, 3053–3058.
23. Kräusslich, H.-G., Nicklin, M. J. H., Toyoda, H., Etchison, D. & Wimmer, E. (1987) *J. Virol.* **61**, 2711–2718.
24. Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M. & Oroszlan, S. (1988) in *Human Retroviruses, Cancer, and AIDS: Approaches to Prevention and Therapy* (Liss, New York), pp. 135–147.
25. Kotler, M., Katz, R. A., Danho, W., Leis, J. & Skalka, A. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4185–4189.
26. Schneider, J. & Kent, S. B. H. (1988) *Cell* **54**, 363–368.
27. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. & Varmus, H. E. (1988) *Nature (London)* **331**, 280–282.
28. Pearl, L. H. & Taylor, W. R. (1987) *Nature (London)* **328**, 482.
29. Aoyagi, T., Morishima, H., Nishizawa, R., Kunimoto, S., Takeuchi, T., Umezawa, H. & Ikezawa, H. (1972) *J. Antibiot.* **25**, 689–694.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1990 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 265, No. 23, Issue of August 15, pp. 13890-13898, 1990
Printed in U.S.A.

Recombinant HIV2 Protease Processes HIV1 Pr53^{gag} and Analogous Junction Peptides *in Vitro**

(Received for publication, April 24, 1990)

Sergio Pichuantes‡, Lilia M. Babé‡§, Philip J. Barr¶, Dianne L. DeCamp‡‖, and Charles S. Craik‡

‡*From the Departments of Pharmaceutical Chemistry and Biochemistry/Biophysics, University of California San Francisco, California 94143 and the* ¶*Chiron Corporation, Emeryville, California 94608*

A synthetic DNA fragment encoding a protease precursor of the human immunodeficiency virus type 2 (HIV2) was cloned and expressed in bacteria and yeast. A recombinant plasmid encoding a hybrid polypeptide consisting of human superoxide dismutase and an HIV2 protease precursor of 113 amino acids was constructed for regulated intracellular expression in bacteria. Induction of this plasmid produced an autoprocessed form of the retroviral enzyme possessing the correct molecular weight. Overexpression and secretion of the protease from yeast was achieved with an expression vector encoding the yeast pheromone α-factor signal/leader sequence fused to a protease precursor of 115 amino acids. Amino-terminal sequence analysis confirmed that the viral enzyme exported from yeast was correctly processed from its precursor by cleavage of the predicted Ala-Pro peptide bond located at the NH_2 terminus of the protease in the *pol* open reading frame. No additional amino acid residues were required at the COOH terminus of the protease for this autoproteolytic event. The HIV2 protease expressed in bacteria and yeast was active in an *in vitro* assay when tested on the HIV1 polyprotein precursor, myristylated Pr53^{gag}. Two synthetic peptides representing junction sequences in the HIV1 gag-pol precursor were used to assay purified HIV2 protease. The enzyme exhibited a k_{cat}/K_M of 23.2 min^{-1} mM^{-1} on the HIV1 matrix-capsid junction peptide and a k_{cat}/K_M of 71.4 min^{-1} mM^{-1} on the protease-reverse transcriptase junction peptide. These rates show that the HIV2 enzyme is efficient at hydrolyzing the HIV1 peptide junctions, revealing the analogous nature of the substrate specificities of the two enzymes.

The human immunodeficiency virus (HIV)[1] (1) is the etiological agent of the acquired immunodeficiency syndrome (AIDS), a disease that has evolved into a worldwide public health problem. In addition to HIV type 1 (HIV1) (2-4) which has been implicated in epidemic AIDS in North America, Europe, and Central Africa (5), another retroviral group designated HIV2 (6) has been identified primarily in AIDS patients from West Africa (5, 7). Nucleotide sequence comparisons along with immunological studies carried out on several isolates of HIV2 (6-10) have revealed that this retrovirus is more closely related to the simian immunodeficiency virus (11, 12) than to HIV1. The gag and pol proteins of HIV2 can be precipitated by antibodies in sera from patients infected with HIV1 revealing a conservation of antigenic determinants between the HIV1 and HIV2 proteins. However, the gag and pol products of HIV2 have less than 60% identity in their amino acid sequences (6) when compared with the equivalent HIV1 proteins. Moreover, variation in the size, amino acid composition, and proteolytic cleavage junctions is also evident when the HIV2 polyproteins are compared with their HIV1 counterparts.

The rapid spread of HIV2 (5) underscores the need to understand the molecular and structural biology of this retrovirus for development of strategies to treat and control AIDS infections. One attractive target in the effort to arrest HIV replication is the viral protease encoded at the 5′ end of the pol gene (13). This enzyme plays an essential role in the viral life cycle by processing the gag and gag/pol polyproteins to the mature structural proteins and enzymes required for virion formation. Recent studies of HIV1 protease have resulted in the elucidation of the three-dimensional structure of the enzyme (14-16). These studies confirm that this retroviral enzyme is a homo-dimeric member of the aspartyl protease family as anticipated by others (17) and permit structure-based design of protease inhibitors that may eventually serve as antiviral agents (18-20). Although the HIV1 and HIV2 proteases presumably play a similar role in viral replication, significant differences exist between the two enzymes. The HIV1 and HIV2 proteases are both 99 amino acids in length but share only 47.5% sequence identity (3, 6) and display unique substrate specificities. Moreover, the HIV2 protease lacks the 2 cysteine amino acid residues found in the HIV1 counterpart, does not react with polyclonal antibodies raised against the HIV1 protease, and is predicted to exhibit a much lower isoelectric point. These differences require that an independent analysis be carried out on the HIV2 protease.

The development of rapid and efficient systems to overproduce a soluble and authentic form of this viral enzyme will facilitate further biochemical and biophysical studies. Furthermore, the availability of *in vitro* assays for the HIV2 protease is essential to address specific questions relating to the mode of action and specificity of the enzyme and its engineered variants. Two recent studies have reported on the chemical synthesis (21) and bacterial expression of the HIV2 protease (22). We have previously reported the expression in

* This work was supported by Grant GM 39552 (to C. S. C.) from the National Institute of Health and by Chiron Corporation (to P. J. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipient of a University of California Task Force AIDS Fellowship F88SF122.

‖ Recipient of National Institutes of Health Fellowship GM 13369.

[1] The abbreviations used are: HIV, human immunodeficiency virus; bp, base pair; hSOD, human superoxide dismutase; IPTG, isopropyl-β-D-thiogalactopyranoside; PMSF, phenylmethylsulfonyl fluoride; PR, protease; PVDF, polyvinylidene difluoride; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; HPLC, high performance liquid chromatography; CAPS, 3-[cyclohexylamino]-1-propanesulfonic acid.

bacteria (23) and yeast (24), purification, and initial characterization of the HIV1 protease. Here we demonstrate that a precursor form of the HIV2 protease will autoprocess in bacteria and yeast to yield a mature and active form of the protease. This protease accurately cleaves *in vitro* the heterologous substrate, HIV1 myristylated $Pr53^{gag}$ as well as synthetic peptides of two junctional sequences of the gag-pol precursor.

MATERIALS AND METHODS

Strains—*Escherichia coli* D1210 (25) was used for bacterial expression of the HIV2 protease. *Saccharomyces cerevisiae* AB110 (24) was used for yeast expression and secretion of the HIV2 protease.

Gene and Plasmid Constructions—A 360-base pair synthetic DNA fragment was constructed that encoded a 112-amino acid precursor of the HIV2 protease from isolate ROD (6). The precursor contained 13 additional amino acids at the NH_2 terminus of the 99-amino acid protease and was constructed from 17 overlapping oligonucleotides. Viral codons were chosen to construct the coding sequence except at Val-10 (GTA→GTG), Thr-12 (ACA→ACC), and Val-71 (GTA→GTC). Codons 10 and 12 were chosen to incorporate a unique *Bst*EII site and codon 71 was chosen to incorporate a unique *Hpa*II site. The endonuclease restriction sites *Xba*I, *Nco*I, and *Bgl*II were included sequentially at the 5′ end and a *Sal*I site was included at the 3′ end of the synthetic gene. Fifteen of the oligonucleotides used in the construction were 45 bases in length and had 22- or 23-base pair overlaps. The 5′- and 3′-terminal oligonucleotides were 27 and 18 bases in length, respectively (Fig. 1*A*). The DNA construction included a stop codon at the COOH terminus of the protease that provided the first G of the *Sal*I site at the 3′ end of the gene. The endonuclease restriction sites located at the 5′ and 3′ ends of the DNA segment facilitated the subsequent cloning of the gene into appropriate plasmids and expression vectors. The *Bst*EII and *Hpa*II sites were introduced to permit dissection of the gene. Oligonucleotides were synthesized using solid-phase phosphoramidite chemistry on an Applied Biosystems 380A DNA synthesizer and were purified as described elsewhere (24). Oligonucleotides were phosphorylated, annealed, and ligated following standard protocols (26). Equimolar concentrations (5 pmol) of each phosphorylated oligonucleotide were mixed together in 2 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 10 mM dithiothreitol and heated to 90 °C for 2 min in a 1.5-ml microcentrifuge tube in a covered water bath. The temperature control of the bath was then shut off and the oligonucleotides allowed to anneal for the 4 h needed for the bath temperature to reach 20 °C. The mixture was then ligated at 20 °C for 20 min using T_4 DNA ligase. The DNA was precipitated with ethanol, dried, and suspended in water. After digestion with the enzymes *Xba*I and *Sal*I, the 360-base pair DNA fragment was purified from a 7% polyacrylamide gel and cloned into the M13 vector, mp19. Nine independent clones were isolated and purified, single-stranded DNA was harvested, and the insert of each isolate was sequenced entirely using the chain-termination method (27). Each isolate contained discrepancies when compared with the expected sequence.

One isolate contained three transitions Val-67 (GTA→ATA), Asp 79 (GAC→AAC), Asn 83 (AAC→AAT) and one deletion Ala 92 (GCC→GC) in the coding region of the mature protease. This isolate also contained an insertion/substitution at Arg 5 (AGA→AAGC). Four isolates were identical to the previous isolate except that the C deletion at codon 92 was not present. Three isolates contained four substitutions in the mature protease, Ala-34 (GCA→GTA), Lys-45 (AAA→AAC), Val-71 (GTC→GTA), and Ala-92 (GCC→GCT). One isolate contained two transitions Asp-79 (GAC→AAC) and Asn-83 (AAC→AAT) in the mature protease and one insertion/substitution at Arg-5 (AGA→AAGC). Although the G to A and C to T transitions can be accounted for by guanine modification during chemical DNA synthesis (28), no general pattern can be discerned to account for the discrepancies which may have resulted from machine failure or recombinant inaccuracies. The isolate with one insertion and two single-base substitutions was chosen for repair. The substitution at position 83 was a silent mutation and was left uncorrected. Two oligonucleotides 5′ GGT GCA GCA AGT CCT CTG TTG GTG GCT CCC 3′ and 5′ TGA TTC GGG TGT CGC CTG TCA TTA T 3′ were used to repair the insertion/substitution at Arg-5 and the substitution at Asp-79, respectively. One round of site-directed mutagenesis using both oligonucleotides and methods previously described (29) was used to obtain the desired gene sequence. The repaired DNA was sequenced completely and used for all subsequent DNA constructions.

For internal localization of the HIV2 protease in bacteria, the plasmid pSOD/HIV2PR113 was constructed. The 347-bp *Nco*I-*Sal*I DNA fragment encoding the HIV2 protease was purified and ligated to the *Nco*I-*Sal*I-digested vector pSODCF2 (30). This vector provided the β-lactamase gene for selection, the *Col*E1 origin of replication for autonomous replication in *E. coli* and the inducible *tac* promoter for transcriptional control. Plasmid pSOD/HIV2PR113 encodes a polypeptide consisting of human superoxide dismutase (hSOD) fused to a 113-amino acid precursor of the HIV2 protease (Fig. 1*B*). The cloning strategy involved replacing the COOH-terminal alanine of hSOD with methionine and glycine 72 and threonine 74 of the pol polyprotein (6) by an alanine and a leucine, respectively. Transformation and selection of recombinants was accomplished using standard procedures (26).

For secretion of the HIV2 protease from yeast, the plasmid pHIV2PR115 was constructed. The 339-bp *Bgl*II-*Sal*I DNA fragment encoding the HIV2 protease was cloned into the pBS24.1-derived plasmid pPR179 (24). This yeast expression vector contains 2μ sequences for autonomous replication in yeast, the glucose-regulated hybrid promoter ADH2/GAPDH, the α-factor terminator to ensure transcription termination, and the yeast genes, *leu2-d* and *ura3* for selection. The β-lactamase gene and the *Col*E1 origin of replication are also present in this shuttle vector. Plasmid pHIV2PR115 encodes an HIV2 protease precursor of 115 amino acids with its NH_2 terminus fused to α-factor signal/leader sequences that contain the KEX2 processing site (31) (Fig. 1*C*). The HIV2 protease precursor expressed from this vector contains the sequence Phe-Arg-Glu-Asp-Leu (residues 70–74 of the pol polyprotein) at its NH_2 terminus instead of the corresponding sequence Ala-Gly-Gly-Asp-Thr (6) due to genetic manipulations required for cloning purposes. Yeast transformation and selection of leucine/uracil prototrophs in leucine- and uracil-deficient media was achieved essentially as previously described (32).

Expression of HIV2 Protease in Bacteria and Yeast—*E. coli* D1210 cells harboring plasmid pSOD/HIV2PR113 were grown for 15–18 h at 37 °C in 3 ml of Luria broth (26) containing ampicillin at 40 μg/ml. The overnight culture was used to inoculate 60 ml of M9 minimal medium containing ampicillin at 40 μg/ml (26) and cells were grown for 2 h at 37 °C with vigorous shaking. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added at 200 μM final concentration, and cultures were incubated for another 18 h. Aliquots of the induced culture containing 0.3 OD_{650} were collected by centrifugation at 10,000 × *g* for 10 min, and pellets were lysed by repeated boiling and freezing in 40 μl of 63 mM Tris-HCl, pH 6.8, 50 mM dithiothreitol, 10% glycerol, 3% sodium dodecyl sulfate (SDS), 0.02% bromphenol blue. Proteins fractionated by SDS-PAGE in 15% polyacrylamide gels were transferred to nitrocellulose filters for 1 h at 75 volts using a Bio-Rad Trans-Blot apparatus. The HIV2 protease was identified on immunoblots using rabbit polyclonal antibodies raised against a synthetic HIV2 protease[2] as the primary antibody. A goat anti-rabbit antibody conjugated to horseradish peroxidase (Tago) was used as the secondary antibody for immunodetection.

Cultures of *S. cerevisiae* AB110 cells harboring pHIV2/PR115 were grown in 50 ml of leucine- and uracil-deficient media for 24–48 h at 30 °C, and cells were collected by centrifugation at 7000 × *g* for 20 min. Proteins in 5 ml of supernatants were precipitated with 2.5 ml of 50% trichloroacetic acid containing 2 mg/ml deoxycholate. After incubation at 4 °C for 30 min, pellets were collected, washed with acetone, dried, and suspended in 20 μl of sample buffer (33). Proteins fractionated by SDS-PAGE were visualized by Coomassie Blue staining or transferred to nitrocellulose filters for immunodetection.

For large scale yeast fermentation and expression of the HIV2 protease, a 115-liter fermentor was prepared with 68 liters of leucine- and uracil-deficient media (32) containing 0.05% (w/v) deoxycholate. The fermentor was then inoculated with a 6-liter seed culture of *S. cerevisiae* AB110 harboring the recombinant plasmid pHIV2PR115. The seed culture had been grown for 24 h in leucine-deficient media to ensure a high copy number of the expression plasmid. Culture conditions were 30 °C, 5.0 psi and 300 rpm. Cell density, pH, CO_2 concentration, and glucose concentration measurements were made every 2 h throughout the duration of the fermentation run, which lasted 72 h. Cell density measurements were made using an in-line spectrophotometer set at 650 nm (OD_{650}). The pH was determined with an in-line pH electrode. The concentration of CO_2 was determined by infrared spectroscopy using an IR-703 gas analyzer (In-

[2] E. Bradley and I. Kuntz, unpublished results.

frared Industries, Inc., Santa Barbara, CA) mounted on the fermentation vessel to sample the head space. Glucose concentration measurements were made using a hexokinase-coupled assay and ultraviolet determination of NADH levels at 340 nm. Expression levels of the HIV2 protease at 24, 30, 48, 52, and 72 h after inoculation were estimated by immunoblotting as described above.

Preparation of Bacterial Protein Extracts and Concentrated Yeast Supernatants—Bacterial pellets (50 OD_{650}) of cultures grown for 4 h after induction, as described above, were suspended in 1 ml of 50 mM Tris-HCl, pH 8.0, 100 mM KCl, 5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 0.5% Triton X-100, and cells were disrupted by sonication for 1 min at 4 °C. Lysates were clarified by centrifugation. All centrifugations were at 45,000 × *g* for 15 min. Protamine sulfate was added to the supernatant to a final concentration of 0.5% (w/v). After incubation at 20 °C for 15 min, the precipitate was removed by centrifugation. Ammonium sulfate was added to the supernatant to 15% of saturation (w/v), incubated at 4 °C for 1 h, and a pellet was collected by centrifugation. Saturated ammonium sulfate was then added to the supernatant to a final concentration of 40% (w/v). After incubation for 1 h at 20 °C, the pellet was collected, resuspended in 0.2 ml of Tris-HCl, pH 7.5, and dialyzed against the same buffer using a microdialyzer (Bethesda Research Laboratories) with membranes of 3,500 molecular weight cut-off.

HIV2 protease expressed and secreted from yeast in volumes less than 100 ml was concentrated by ultrafiltration from the media of yeast AB110 cells harboring pHIV2PR115. Cultures grown for 48 h were centrifuged to remove the cells, and the supernatant was concentrated approximately 20-fold at 4 °C, using a Centricon 10 (Amicon) membrane with a molecular weight cut-off of 10,000.

Protein Purification—The HIV2 protease was purified from the media of yeast cultures expressing plasmid pHIV2PR115 grown 72–96 h. During the initial stages of purification, the protease was monitored by immunoblots using polyclonal rabbit antibodies raised to a chemically synthesized HIV2 99 amino acid peptide. Following centrifugation of the cells, the culture supernatant (5 liters) was adjusted to 1 M ammonium sulfate, 20 mM sodium phosphate, pH 8.0, 1 mM EDTA and 1 mM phenylmethylsulfonyl fluoride. This material was loaded on a phenyl-Sepharose column (5 × 30-cm, Pharmacia LKB Biotechnology Inc.) pre-equilibrated with the same buffer. The column was washed with a linear gradient of 1–0 M ammonium sulfate in buffer followed by two column volumes of water. The protease, which eluted in the later fractions of the water wash in approximately 250 ml, was further purified by preparative isoelectric focusing on a Rotofor unit (Bio-Rad). A mixture of 1 ml, pH 3–10, Ampholytes and 0.25 ml, pH 5–8, Ampholytes was added to 55 ml of protease sample for each of five Rotofor fractionations. The fractions in the range of pH 5.0–6.5 were pooled. These samples were further purified by reversed-phase HLPC using a preparative C_3 column (Pro 10/300 Protein Plus, Du Pont, 4.6 × 21.2 cm) and eluted with a linear gradient of 25–85% acetonitrile/water in 0.1% trifluoroacetic acid over a period of 60 min at a flow rate of 10 ml/min. The protein eluted at approximately 60% acetonitrile and was lyophilized and stored at −20 °C. The resulting enzyme is 95% pure as judged by NH_2-terminal sequence analysis as well as silver staining of SDS-PAGE gels. For enzymatic analysis, the lyophilized enzyme was denatured in 0.05 mM sodium acetate, pH 5.5, containing 8 M urea, and refolded by a 10-fold dilution into 0.2 M sodium acetate, pH 5.5, containing 1 mM EDTA, 5 mM dithiothreitol, 10% glycerol, and 5% ethylene glycol. This procedure has been shown to restore enzymatic activity to preparations of HIV1 enzyme that have become inactive due to denaturation (34).

In Vitro Assay for HIV2 Protease on $Pr53^{gag}$—A recombinant form of the natural substrate $Pr53^{gag}$ has previously been expressed in yeast (35). The myristylated polyprotein was purified and shown to be a substrate for the HIV1 protease (24). In a similar fashion, we monitored the activity of the HIV2 protease expressed in bacteria and yeast on myristylated $Pr53^{gag}$. The heterologous substrate (1 μg) was incubated in 10 mM Tris-HCl, pH 7.0, containing 130 mM NaCl, 1 mM EDTA, and 1 mM phenylmethylsulfonyl fluoride with the concentrated yeast media or partially purified bacterial lysates. After 8 h of incubation at 25 °C, the reaction products were analyzed by SDS-PAGE in 15% polyacrylamide gels and immunoblotted using serum from AIDS patients as the primary antibody. The serum was inactivated in a Biosafety Level 3 facility by heating at 56 °C for 35 min, treating with psoralen at 25 μg/ml final concentration, and UV irradiation on ice. Visualization of the specific bands on the immunoblots was achieved with goat anti-human antibodies conjugated to horseradish peroxidase (Tago) as the second antibody. Lysates of HIV1-infected cells (36) were included as markers to visualize HIV1 viral proteins. The lysates were inactivated in a Biosafety Level 3 facility by treating with 0.5% Triton X-100 final concentration. Pepstatin A was used at 10 mM final concentration in some assays to inhibit the activity of the retroviral protease.

In Vitro Assay for HIV2 Protease on Synthetic Peptide Substrates—The HIV2 protease was assayed against the decapeptide, Ala-Thr-Leu-Asn-Phe-Pro-Ile-Ser-Pro-Trp and the octapeptide, Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln. The decapeptide corresponds to the HIV1 carboxyl-terminal autoprocessing site, and the octapeptide corresponds to the HIV1 matrix-capsid cleavage site (2). The peptides were synthesized using conventional solid-phase methods. Concentrations of the enzyme stock solutions were established by titration with the substrate-based inhibitor Val-Ser-Gln-Asn-LeuΨ($CH(OH)CH_2$)Val-Ile-Val (34). Reactions were carried out in 0.1 ml of 0.1 M sodium acetate buffer, pH 4.7, containing 2 mM EDTA, 1 M NaCl, and 0.05–5 mM peptide substrate. The effect of salt and pH on enzyme activity was tested by using 0.1 M sodium acetate buffer at either pH 4.7 or 5.5 and containing either 0.25 or 1 M NaCl. Typically, $1–3 \times 10^{-4}$ mg of HIV2 protease was added to initiate the reaction, which was stopped after 2 h at 37 °C by addition of 50 μl of cold 0.1% trifluoroacetic acid on ice. Conditions were adjusted so that <25% of the substrate was hydrolyzed during the incubation. Reaction products were separated by reversed-phase HPLC using a C_{18} Pecosphere (0.46 × 3.3 cm, Perkin-Elmer) column. Products of the decapeptide (Ala-Thr-Leu-Asn-Phe and Pro-Ile-Ser-Pro-Trp) were resolved with a gradient of 10–50% acetonitrile in 0.1% trifluoroacetic acid, while the octapeptide fragments (Ser-Gln-Asn-Tyr and Pro-Ile-Val-Gln) were observed with a gradient of 0–35% acetonitrile. Absorbance was monitored at 280 nm, and hydrolysis of the peptides was quantitated by integration of the peak areas and comparison to product standard curves. Each data point was measured in triplicate. The program Enzfitter was used to fit data to Michaelis-Menten kinetics (37).

Amino-terminal Sequence Analysis—The retroviral protease in 20 ml of yeast supernatant was obtained from yeast AB110 cultures expressing the recombinant plasmid pHIV2PR115. Proteins were precipitated with trichloroacetic acid, fractionated in a 15% polyacrylamide gel containing SDS and transferred electrophoretically to polyvinylidene difluoride membranes (Immobilon-P Transfer Membrane, Millipore) in 10 mM CAPS, pH 11.0, 10% methanol. Strips of polyvinylidene difluoride membranes were excised and the protease subjected to microsequence analysis. Automated Edman degradation chemistry was performed using a 470A Applied Biosystems gas-phase sequencer equipped with an on-line phenylthiohydantoin amino acid analyzer. Phenylthiohydantoin-amino acids were separated by reversed-phase chromatography on a Brownlee C_{18} column and identified as described previously (38).

RESULTS

Bacterial expression of the HIV2 protease was achieved by cloning a synthetic DNA fragment (Fig. 1*A*) encoding the retroviral enzyme into a bacterial expression vector derived from pSOD/PR179 (23). The resulting recombinant plasmid pSOD/HIV2PR113 (Fig. 1*B*) encodes a polyprotein containing a 153-amino acid human superoxide dismutase and an HIV2 protease precursor of 113 amino acids. The hSOD sequences are essential for expression of the 113-amino acid protease precursor since a similar construction lacking the hSOD sequences resulted in no detectable HIV2 protease. The viral protease precursor contains an NH_2-terminal extension of 14 amino acids to the 99-amino acid protease. There is no COOH-terminal extension since a stop codon was placed immediately downstream of the COOH-terminal leucine of the protease. The expression of the fusion protein was controlled by IPTG induction of the *tac* promoter located 5′ to the hSOD gene. Protease expression was monitored in total cell lysates by SDS-PAGE and immunoblotting. An immunoreactive band of approximately 10 kDa (PR) can be observed within 0.25 to 0.5 h of induction using rabbit polyclonal antibodies raised against a synthetic HIV2 protease (Fig. 2, *lanes 2* and *3*). The 10-kDa band was not detected using rabbit polyclonal antibodies raised against the HIV1 protease. The 10-kDa species is in agreement with the 10,721-dalton pre-

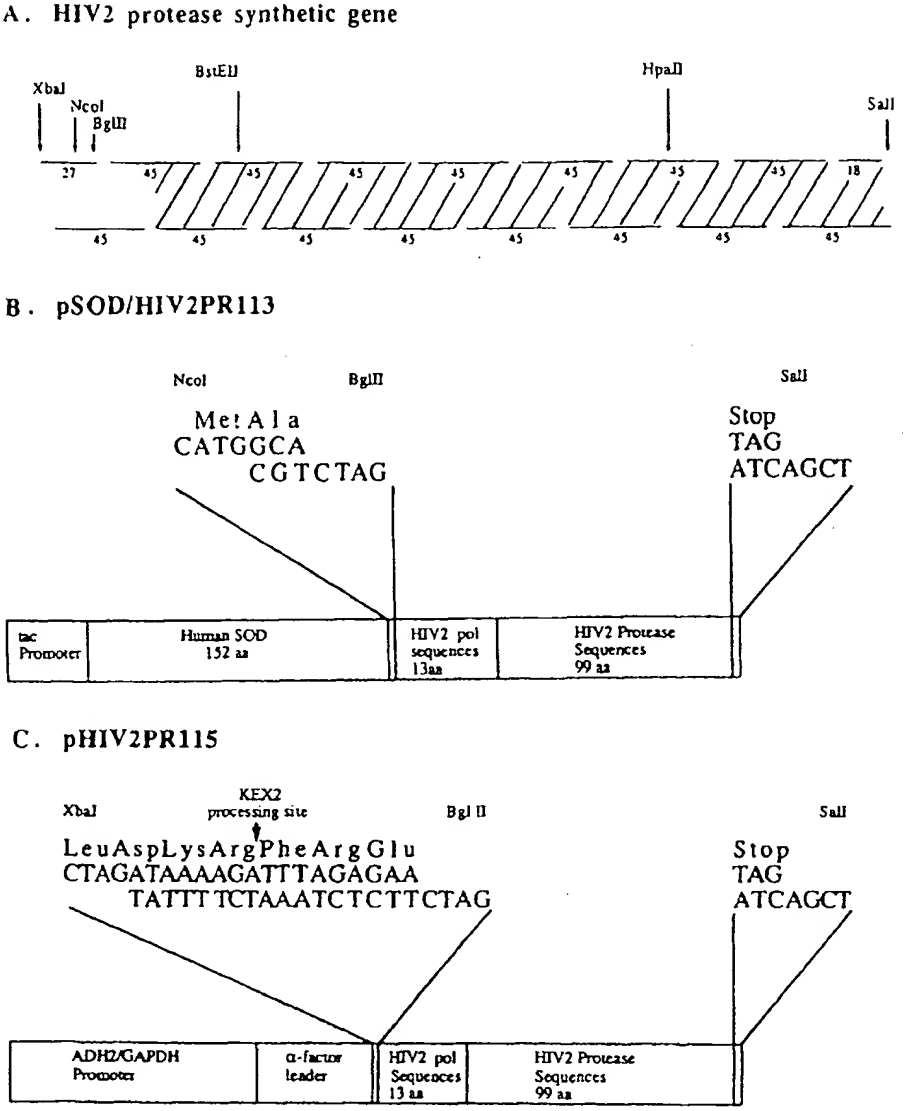


FIG. 1. **Gene construction and recombinant plasmids for the expression of the HIV2 protease.** *A*, synthetic HIV2 protease gene. Schematic representation of the HIV2 protease synthetic gene. A 360-bp *XbaI-SalI* DNA fragment was constructed using 17 overlapping synthetic oligonucleotides of 18, 27, and 45 nucleotides in length. The *XbaI*, *BglII*, and *SalI* restriction sites were included to facilitate subcloning of the final construction. The *BstEII* and *HpaII* restriction sites present in the coding region of the 99-amino acid protease (*hatched*) were introduced to permit dissection of the gene. *B*, bacterial expression plasmid pSOD/HIV2 PR113. A 339-bp synthetic *BglII-SalI* DNA fragment encoding the 99-amino acids HIV2 protease and containing 13 additional residues at its NH_2 terminus was cloned into the plasmid pSOD/PR179 (23). The resulting plasmid encodes a hSOD/HIV2 protease hybrid precursor whose expression is under the control of the *tac* promoter. *C*, yeast expression plasmid pHIV2PR115. The *BglII-SalI* restriction fragment described above was cloned into the plasmid pPR179 (24) that provided the yeast α-factor signal/leader sequence containing the KEX-2 processing site to promote secretion of the retroviral enzyme from the cell. The HIV2 recombinant plasmid also contains the ADH2/GAPDH promoter, the α-factor terminator, 2μ yeast sequences and the genetic markers *leu2-d*, *ura3*, and β-lactamase.

dicted mass of the mature protease. This indicates that the HIV2 protease is capable of autoprocessing within the bacterial host as previously reported for the HIV1 protease (23, 39-41). Another major band of approximately 30 kDa (SOD/PR) was also detected by the antibodies to the HIV2 protease. This band is also detected by a mouse monoclonal antibody raised against hSOD (results not shown), and probably represents the 27,865-kDa SOD/HIV2 protease fusion containing the unprocessed viral enzyme. As expected, the protease and the hybrid precursor were not observed in the uninduced cultures.

Expression levels of the hybrid polypeptide and the mature protease reach their maximum levels at approximately 3–4- and 5–6 h post-induction, respectively (Fig. 2, *lanes 8–11*). Expression levels of both proteins are greatly reduced after 18 h suggesting that these proteins are susceptible to degradation by *E. coli* proteases. Increasing levels of the 30-kDa precursor band do not result in a concomitant increase of the 10-kDa protease band suggesting that release of the protease from the hybrid polypeptide is slow relative to the proteolytic turnover of the viral protease. The HIV2 protease is completely soluble in this expression system and at maximum expression levels is approximately 0.1% of the total cellular proteins.

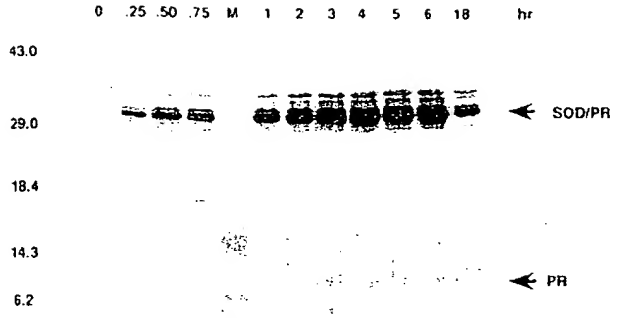


FIG. 2. **Immunodetection of HIV2 protease expressed in *E. coli* D1210.** Cells harboring plasmid pSOD/HIV2PR113 were grown in 60 ml of M9 minimal media (26) for 2 h, and cultures were induced by the addition of IPTG at a final concentration of 0.2 mM. Samples were removed at different times post-induction and 0.3 OD_{600} of cells were lysed as described under "Materials and Methods." Proteins fractionated by electrophoresis in a 15% polyacrylamide gel containing SDS were transferred to nitrocellulose filters. Immunoblot analysis was performed using rabbit polyclonal sera raised against a synthetic HIV2 protease. The column numbers indicate hours after induction with IPTG. *PR* is the protease; *SOD/PR* is the hybrid precursor. Prestained proteins were loaded as molecular weight standards (*column M*).

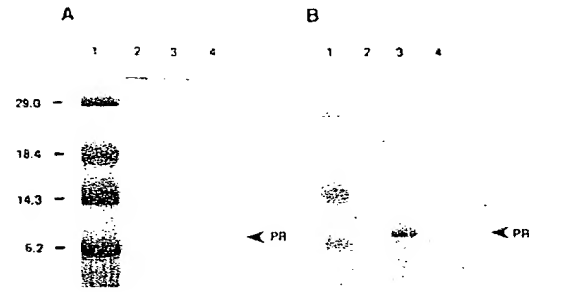


FIG. 3. **Detection of HIV2 protease in yeast supernatants of *S. cerevisiae* AB110.** *A*, Coomassie Blue-stained proteins. *B*, immunoreactivity with rabbit sera containing antibodies to a synthetic HIV2 protease. Cells harboring pHIV2PR115 were grown in 50 ml of leucine and uracil deficient media at 30 °C. After 24 (*lane 2*) or 48 h (*lane 3*) of incubation, 10-ml aliquots of the cultures were pelleted by centrifugation and proteins in 5 ml of yeast supernatants were precipitated with trichloroacetic acid and fractionated by SDS-PAGE in a 15% polyacrylamide gel containing SDS. Immunoblot analysis was performed as described in Fig. 2. Trichloroacetic acid-precipitated proteins of yeast supernatants of AB110 cells harboring the parental plasmid pBS24.1 (24) are shown in *lane 4*. *PR* is the mature HIV2 protease.

To avoid the problems associated with intracellular proteolysis in bacteria as well as to facilitate its purification, the protease was expressed and secreted in yeast. Yeast expression of the HIV2 protease was accomplished by fusing the synthetic DNA fragment encoding the HIV2 protease to DNA encoding the α-factor signal/leader sequences that contain the KEX2 recognition cleavage site (Fig. 1*C*). The glucose-regulatable hybrid promoter ADH2/GAPDH (42) is used to control expression of the cloned gene. The α-factor terminator (43) is included to ensure transcription termination. The plasmid pHIV2PR115 encodes an HIV2 pol precursor of 115 amino acids with a predicted molecular weight of 12,435. A protein band of approximately 10 kDa (PR) is observed by Coomassie Blue staining (Fig. 3*A*, *lane 3*) and immunoblotting (Fig. 3*B*, *lane 3*) in yeast supernatants of cells transformed with pHIV2PR115, indicating autoprocessing of the precursor. The viral protease is not observed in yeast supernatants of cells grown for 24 h (*A* and *B*, *lane 2*) since glucose consumption and the resultant induction of the ADH2/GAPDH promoter takes place after 28–30 h of growth under the culture conditions used (see below). Maximum levels of expression (0.8–2.0 mg/liter of media) are observed after 48–72 h. The HIV2 protease proved to be stable at 30 °C in the yeast media for up to 6 days as judged by immunoblotting and enzymatic analyses (results not shown) and represents a predominant yeast-secreted protein. This protein is not seen in yeast supernatants of cells harboring the parental plasmid pBS24.1 (*A* and *B*, *lane 4*).

The NH_2 terminus of the yeast-secreted HIV2 protease was sequenced to unequivocally demonstrate that the 10-kDa product represents the correctly processed, mature, retroviral protease. The first 13-amino acid residues displayed the sequence Pro-Gln-Phe-Ser-Leu-Trp-Lys-Arg-Pro-Val-Val-Thr-Ala which is identical to that predicted for the HIV2 protease (6). This confirms that the viral enzyme secreted from yeast was correctly released from the precursor by cleavage at the Ala-Pro junction.

To determine whether the recombinant HIV2 protease expressed in bacteria and yeast represents an active form of the viral enzyme, its ability to correctly cleave the HIV1 polyprotein substrate, myristylated $Pr53^{gag}$, was established. Both the bacterial- (Fig. 4, *lane 2*) and yeast- (*lane 6*) expressed protease can accurately process the HIV1 protein precursor as judged by the generation of capsid protein (*p24*) and matrix protein (*p17*) along the course of the incubation. Another protein band of approximately 6 kDa (*p6* or *p7*) that probably represents a processing product of nucleocapsid protein (*p15*) (44) was also detected by the human AIDS serum. This protein had remained undetected when similar experiments were performed with the HIV1 protease (24) due probably to a low titer of antibodies against this protein. These processed species are indistinguishable from those generated by a purified HIV1 protease expressed in yeast (*lane 3*) (24) and comigrate with their counterparts expressed *in vivo* (*lane 4*). Bacterial extracts and yeast supernatants that do not contain the HIV2 protease do not generate specific viral proteins when incu-

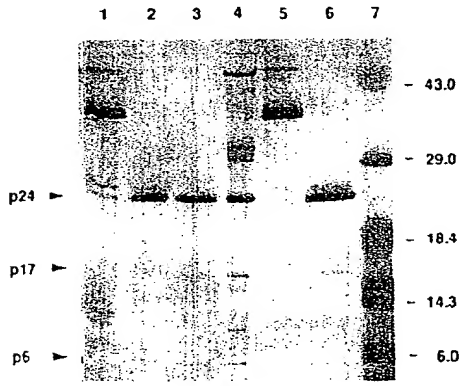


FIG. 4. **Immunoblot analysis of *in vitro* processing of HIV1 $Pr53^{gag}$ polyprotein by bacterial and yeast expressed HIV2 protease.** For each *in vitro* assay, approximately 1 μg of myristylated $Pr53^{gag}$ precursor was incubated at 25 °C with bacterial extracts or concentrated yeast supernatants as described under "Materials and Methods." Immunoblot analysis was performed using AIDS patient sera as the primary antibody. *Lane 1*, $Pr53^{gag}$ incubated with bacterial extracts of cells harboring pSOD/HIV2PR113 before induction; *lane 2*, $Pr53^{gag}$ incubated with bacterial extracts of cells harboring pSOD/HIV2PR113 4 h after induction; *lane 3*, $Pr53^{gag}$ incubated with purified HIV1 protease expressed in yeast (24); *lane 4*, a total cell lysate of HIV1 infected cells (36) as reference for viral proteins; *lane 5*, $Pr53^{gag}$ incubated with concentrated yeast supernatants from yeast cells harboring pBS24.1; *lane 6*, $Pr53^{gag}$ incubated with concentrated yeast supernatants from yeast cells harboring pHIV2PR115; *lane 7*, prestained molecular weight markers.

bated with $Pr53^{gag}$. Nonspecific proteolysis of the HIV1 polyprotein precursor (*lanes 1* and *5*) is presumably caused by endogenous bacterial and yeast proteases. Processing of $Pr53^{gag}$ by the HIV2 protease was completely abolished when 10 mM pepstatin was included in the assay (results not shown). This observation provides evidence that the HIV2 protease is an aspartyl protease as already established for the HIV1 enzyme (15).

The HIV2 protease was purified from yeast supernatants using a three-step purification procedure. Fig. 5 shows representative samples from the various stages of the protocol. We estimate that the initial concentration of the protease is approximately 0.8 mg/liter of yeast supernatant (*lane 2*). The subsequent steps involving phenyl-Sepharose chromatography, preparative isoelectric focusing fractionation, and reversed-phase HPLC result in homogeneous enzyme (*lanes 3–5*). The phenyl-Sepharose column, which mimics the hydrophobic peptide substrate, is used to concentrate the enzyme from the large volume of yeast media. The protease-containing fractions eluted from the column are subjected to preparative isoelectric focusing on a Rotofor unit to remove high molecular weight proteins and melanin polymers of heterogeneous size. Migration of the HIV2 protease to its expected isoelectric point of approximately 5.5 separates the enzyme from the negatively charged polymeric contaminants which migrate as a dark brown band in the pH range of 1–3. Following removal of the Ampholytes by dialysis, the pH 5.0–6.5 Rotophor fractions are loaded onto a preparative C_3 column and the protease is eluted as a single peak at approximately 60% acetonitrile; this is in agreement with its expected hydrophobic nature. The single band shown in Fig. 5, *lane 5*, is 95% pure as judged by amino acid analysis and silver staining (data not shown). The overall yield of the protein is between 10 and 20%. The purified protein is stable to rapid freeze/thaw cycles and storage at −20 °C in 10% glycerol, 5% ethylene glycol buffer. However, concentrated enzyme solutions (0.25 mg/ml) kept at 0 °C can lose activity, presumably due to autoproteolysis. For cases where enzyme activity is lost due to denaturation, activity can be restored by refolding the enzyme using a procedure similar to that used for HIV1 protease (34). The purified enzyme was assayed against the HIV1 matrix-capsid octapeptide Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln to yield a specific activity of 3.0 μmol/min/mg. This value is comparable to that obtained for hydrolysis of the same substrate by purified HIV1 protease (45).

The kinetic properties of the purified HIV2 protease were evaluated using the two synthetic peptide substrates Ala-Thr-Leu-Asn-Phe-Pro-Ile-Ser-Pro-Trp and Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln. To ensure accuracy in all enzymatic analyses, active site titrations using the substrate-based inhibitor Val-Ser-Gln-Asn-LeuΨ(CH(OH)CH_2)Val-Ile-Val (34) were used to determine the concentration of enzyme stock solutions. Evaluation of two buffer conditions commonly used for HIV1 protease assays showed that an 80% increase in enzyme activity on either peptide substrate resulted by increasing the salt concentration from 25 mM to 1.0 M. A further increase in enzyme activity of approximately 10% resulted by lowering the pH from 5.5 to 4.7. The optimal conditions of high salt and low pH were used for determining the kinetic constants of the octapeptide and decapeptide substrates. The enzyme exhibited a k_{cat} of 65 min^{-1} and a K_M of 2.8 ± 0.5 mM on the octapeptide substrate and a k_{cat} of 10 min^{-1} and a K_M of 0.140 ± 0.045 mM on the decapeptide substrate (Fig. 6).

The behavior of yeast expressing the HIV2 protease in large scale fermentation conditions was evaluated to establish baseline growth and production profiles. Sixty-eight liters of pHIV2PR115 transformed yeast were grown in a 115-liter vessel. Results shown in Fig. 7*A* reveal that the yeast culture remains in lag phase for approximately 15 h until an increase in the cell density and a simultaneous decrease in the pH is observed. The stationary phase is reached after approximately 65 h. Glucose utilization by the yeast resulted in a sharp evolution of CO_2 production from 20 to 28 h post-inoculation (*B*). A concomitant shift in pH was also noted during this time interval (*A*). As expected, the HIV2 protease was expressed upon depletion of available glucose and was detected immunologically after 30 h (*C, lane 4*). Maximal expression

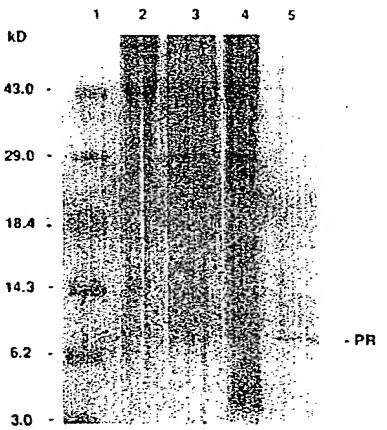


FIG. 5. **Purification of secreted HIV2 protease from yeast.** Samples from various steps of the purification procedure were run on a 17.5% polyacrylamide gel containing SDS and stained with Coomassie Blue. Samples were precipitated with trichloroacetic acid/deoxycholate as described under "Materials and Methods" to reduce the volume of the sample. *Lane 1*, molecular weight markers; *lane 2*, yeast culture supernatant (5 ml, 0.1% of total); *lane 3*, pooled, protease-containing fractions from the phenyl-Sepharose column (0.2% of total); *lane 4*, preparative isoelectric focusing fractions migrating in the pH 5.0–6.5 range (0.6% of total); *lane 5*, purified protease (3.5 μg, 0.8% of total) isolated by reversed-phase chromatography on a preparative C_3 HPLC column.

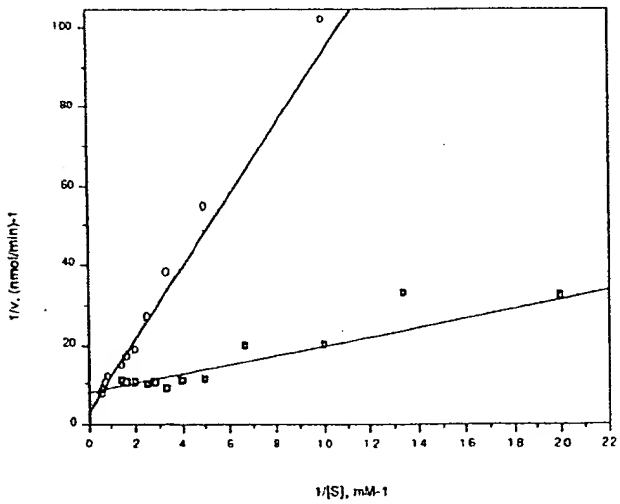


FIG. 6. **Double-reciprocal plot of HIV-2 protease activity *versus* concentration of peptide substrates ATLNFPISPW (■) and SQNYPIVQ (○).** Either 2.5×10^{-4} mg (■) or 9.9×10^{-5} mg (○) HIV-2 protease was incubated with varying amounts of substrate for 2 h at 37 °C in 0.1 M sodium acetate buffer, pH 4.7, containing 1 M NaCl and 2 mM EDTA. Enzyme activity was measured as described under "Materials and Methods." Data were fit to the Michaelis-Menten equation and kinetic constants were calculated using a non-linear regression program ("Enzfitter" from Biosoft).

The Journal of Biological Chemistry

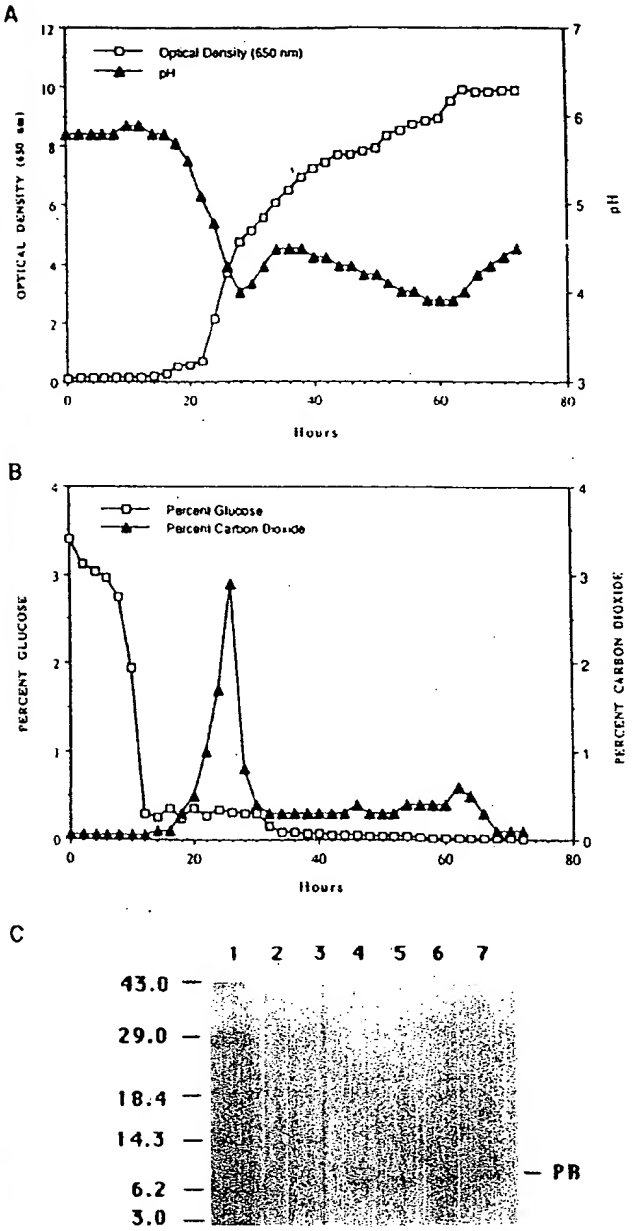


FIG. 7. Fermentation profiles of yeast expressing and secreting HIV2 protease. *A*, cell density (OD_{650}) and pH profiles. *B*, percent glucose concentration and percent CO_2 concentration. *C*, expression levels of secreted HIV2 protease from yeast. The HIV2 protease in 4 ml of yeast supernatant was precipitated with trichloroacetic acid/deoxycholate, fractionated by electrophoresis in a 15% polyacrylamide gel containing SDS and transferred to nitrocellulose filters for immunoblot analysis as described under "Materials and Methods." Protein molecular weight markers, *lane 1*; media from pBS24.1 transformed yeast, 72 h, *lane 2*; media from pHIV2PR115 transformed yeast, 24 h, *lane 3*; 30 h, *lane 4*; 48 h, *lane 5*; 52 h, *lane 6*; 72 h, *lane 7*.

levels were obtained at 72 h (*C*, *lane 7*) and the growth was terminated due to the increasing pH of the media.

DISCUSSION

The emergence and rapid spread of HIV2 as another causative agent of AIDS (5, 6) raises the challenge to better understand the molecular biology of this retrovirus. Since the protease is essential for viral replication and represents an attractive target for anti-AIDS therapeutics we have cloned and expressed the HIV2 protease in bacteria and yeast. This should facilitate the purification of reagent quantities of protein for further biochemical and biophysical studies. In addition, the expression systems allow us to study the autoprocessing event as well as the proteolytic activity of the mature protease on the heterologous substrate HIV1 $Pr53^{gag}$.

The HIV2 protease is capable of autoprocessing in bacteria to form a mature and active form of the enzyme when expressed as part of a human superoxide dismutase fusion protein. The protease in the hSOD/HIV2 polyprotein has only a 14-amino acid NH_2-terminal extension and does not contain additional amino acids at its COOH terminus. The mature HIV2 protease was completely soluble, and the onset of degradation could be monitored after 4 h of induction. The autoprocessing event appears to be less efficient than that observed for the hSOD/HIV1 protease precursor (23) since a large percentage of the HIV2 protease remains in the hybrid polypeptide throughout an 18-h time course of expression. Although the lack of a COOH-terminal extension did not prevent autoprocessing of HIV2, it may have contributed to the inefficiency of the process. Analogous amino acid sequences to those found in the hSOD/HIV1 polypeptide (23) may be required in the hSOD/HIV2 polypeptide for efficient autocatalysis from the polyprotein precursor.

When expressed and secreted in yeast, the HIV2 protease was correctly released from a 115-amino acid pol precursor lacking a COOH-terminal extension. This observation can also be extended to the HIV1 protease as well since a 154-amino acid HIV1 pol precursor that does not contain additional amino acid sequences at the COOH-terminus of the protease efficiently self-processes in yeast.[3] It has been postulated that the HIV protease undergoes autocatalytic activation when expressed as part of a higher molecular weight precursor and that this activation is initiated at the COOH terminus of the protease (40). The mature HIV1 and HIV2 proteases that we have obtained using yeast expression plasmids encoding HIV precursors with no additional amino acid sequence at their COOH termini, support this idea and shows that the autocatalytic event does not require that the COOH-terminal amino acids be present for processing the NH_2-terminal site.

The HIV2 protease monomer is predicted to be a 99-amino acid protein that is released from the gag/pol polyprotein after cleavage of NH_2-terminal Ala-Pro and COOH-terminal Leu-Pro peptide bonds. The amino-terminal sequence that we determined for the yeast-expressed HIV2 protease represents the first direct evidence that the potential Ala-Pro cleavage site is recognized and hydrolyzed by the HIV2 enzyme during autocatalysis. Autoprocessing of the HIV1 protease at the analogous Phe-Pro cleavage sites has been reported (23, 24, 39–41). Mass spectral analysis of purified HIV2 protease provides additional evidence that the mature HIV2 protease is correctly processed to form the 99-amino acid enzyme monomer.[4]

To determine whether the bacterial and yeast-expressed HIV2 protease represents an active viral product with authentic enzymatic activity, the pattern of proteolysis using purified HIV1 $Pr53^{gag}$ precursor as a substrate (35) was determined. Processing of murine leukemia virus and feline leukemia virus gag precursors by a heterologous retroviral protease has been reported (46–48). Both the bacterial and yeast-expressed

[3] S. Pichuantes, L. M. Babé, P. J. Barr, D. L. DeCamp, and C. S. Craik, unpublished results.

[4] S. Kaur, L. M. Babé, D. L. DeCamp, A. Burlingame, and C. S. Craik, unpublished results.

HIV2 protease process the HIV1 gag polyprotein to yield a product pattern that is indistinguishable from that generated by an HIV1 protease expressed in yeast. The migration of the proteins generated *in vitro* is also identical to the migration of structural proteins from HIV1 virions. This shows that the HIV2 protease expressed in both the bacterial and yeast systems are enzymatically active and capable of authentic processing of the HIV1 gag precursor *in vitro*.

The HIV2 gag polyprotein has a calculated molecular weight of 57,100 and is 20 amino acids larger than its HIV1 counterpart (3, 6). The HIV1 and HIV2 gag precursors share 58% identity in their amino acid sequences. Processing of the HIV1 $Pr53^{gag}$ by the HIV2 protease suggests that the HIV1 and HIV2 gag polyproteins adopt similar conformations that allow the recognition and hydrolysis of a select number of peptide bonds within the polyprotein precursor. Synthetic peptides with sequences of the HIV1 matrix-capsid junction and the protease-reverse transcriptase junction were shown to be substrates for the HIV2 protease in quantitative assays using purified protease whose concentration was verified by active site titration. This permits an accurate comparison of the specificities of the HIV1 and HIV2 enzymes on the same substrate. The HIV2 protease hydrolyzes the octapeptide substrate with a k_{cat} of 65 min^{-1} and a K_M of 2.8 mM. The similarity of these kinetic parameters to those described for the HIV1 protease on the identical substrate (k_{cat} = 78 min^{-1}, K_M = 1.5 mM) (45) shows the analogous substrate specificity of the two enzymes and confirms the observation of efficient HIV2 protease processing of $Pr53^{gag}$, *in vitro*. A detailed analysis of the HIV1 and HIV2 substrate specificities on various synthetic peptides is described elsewhere (49). These results suggest that active site-directed inhibitors currently being developed against the HIV1 protease may also serve as effective inhibitors of the HIV2 protease. Preliminary results using the purified HIV2 protease and synthetic inhibitors support this proposal (49).

The expression systems described here are shown to be useful in the production of a mature, active HIV2 protease. These systems may also prove to be useful for the expression of variant forms of the HIV2 protease as well as related aspartyl proteases. The large scale production of active HIV2 protease achieved by using the yeast expression system will facilitate its purification as well as permit the development of sensitive and efficient *in vitro* assays for this protein. The availability of reagent levels of homogeneous material will also permit biophysical analysis of the retroviral enzyme to provide a better understanding of structure/function relationships and aid in the rational design of inhibitors and eventual antiviral pharmaceuticals.

Acknowledgments—We gratefully acknowledge the contributions of Dr. F. Masiarz (Chiron Corp.) for gas-phase sequence analysis, Dr. I. Bathurst (Chiron Corp.) for purified $Pr53^{gag}$, Dr. A. Tomasselli (The Upjohn Co.) for the HIV2 protease active site titrant and Dr. D. Pliura (Syntex, Inc) for helpful discussions concerning the protease assay. We also thank Dr. S. Hughes of Advanced Bioscience Laboratories and J. Ross of Program Resources Inc., Fredrick Cancer Research and Development Center for large scale fermentation of yeast.

REFERENCES

1. Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and Weiss, R. (1986) *Science* **232**, 697
2. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., and Wong-Staal, F. (1985) *Nature* **313**, 277–284
3. Sánchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., and Luciw, P. A. (1985) *Science* **227**, 484–492
4. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985) *Cell* **40**, 9–17
5. Piot, P., Plummer, F. A., Mhalu, F. S., Lamboray, J.-L., Chin, J., and Mann, J. M. (1988) *Science* **239**, 573–579
6. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., and Alizon, M. (1987) *Nature* **326**, 662–669
7. Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M.-A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L., and Montagnier, L. (1986) *Science* **233**, 343–346
8. Evans, L. A., Moreau, J., Odehouri, K., Legg, H., Barboza, A., Chen-Mayer, C., and Levy, J. A. (1988) *Science* **240**, 1522–1525
9. Kong, L. I., Lee, S.-W., Kappes, J. C., Parkin, J. S., Decker, D., Hoxie, J. A., Hahn, B. H., and Shaw, G. M. (1988) *Science* **240**, 1525–1529
10. Zagury, J. F., Franchini, G., Reitz, M., Collalti, E., Starcich, B., Hall, L., Fargnoli, K., Jagodzinski, L., Guo, H.-G., Laure, F., Arya, S. K., Josephs, S., Zagury, D., Wong-Staal, F., and Gallo, R. C. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5941–5945
11. Franchini, G., Gurgo, C., Guo, H.-G., Gallo, R. C., Collalti, E., Fargnoli, K. A., Hall, L. F., Wong-Staal, F., and Reitz Jr., M. S. (1987) *Nature* **328**, 539–543
12. Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P., and Sonigo, P. (1987) *Nature* **328**, 543–547
13. Kräusslich, H.-G., and Wimmer, E. (1988) *Annu. Rev. Biochem.* **57**, 701–754
14. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) *Nature* **337**, 615–620
15. Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989) *Science* **245**, 616–621
16. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G., and Hobart, P. M. (1989) *Nature* **342**, 299–302
17. Pearl, L. H., and Taylor, W. R. (1987) *Nature* **329**, 351–354
18. Dreyer, G. B., Metcalf, B. W., Tomaszek, T. A., Jr., Carr, T. J., Chandler, A. C., III, Hyland, L., Fakhoury, S. A., Magaard, V. W., Moore, M. L., Strickler, J. E., Debouck, C., and Meek, T. D. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9752–9756
19. Meek, T. D., Lambert, D. M., Dreyer, G. B., Carr, T. J., Tomaszek, T. A., Jr., Moore, M. L., Strickler, J. E., Debouck, C., Hyland, L. J., Matthews, T. J., Metcalf, B. W., and Petteway, S. R. (1990) *Nature* **343**, 90–92
20. McQuade, T. J., Tomasselli, A. G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K., Heinrikson, R. L., and Tarpley, W. G. (1990) *Science* **247**, 454–456
21. Copeland, T. D., and Oroszlan, S. (1988) *Gene Anal. Tech.* **5**, 109–115
22. Le Grice, S. F. J., Ette, R., Mills, J., and Mous, J. (1989) *J. Biol. Chem.* **264**, 14902–14908
23. Babé, L. M., Pichuantes, S., Barr, P. J., Bathurst, I., Masiarz, F. R., and Craik, C. S. (1990) *UCLA Symp. Mol. Cell. Biol. New Series* **110**, 71–88
24. Pichuantes, S., Babé, L. M., Barr, P. J., and Craik, C. S. (1989) *Proteins Struct. Funct. Genet.* **6**, 324–337
25. Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980) *Gene (Amst.)* **8**, 279–300
26. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
27. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
28. Eadie, J. S., and Davidson, D. S. (1987) *Nucleic Acids Res.* **15**, 8333–8349
29. Evnin, L., and Craik, C. S. (1988) *Ann. N. Y. Acad. Sci.* **542**, 61–74
30. Steimer, K. S., Higgins, K. W., Powers, M. A., Stephans, J. C., Gyenes, A., George-Nascimento, C., Luciw, P. A., Barr, P. J.,

Hallewell, R. A., and Sánchez-Pescador, R. (1986) *J. Virol.* **58**, 9–16
31. Achstetter, T., and Wolf, D. H. (1985) *EMBO J.* **4**, 173–177
32. Barr, P. J., Cousens, L. S., Lee-Ng, C. T., Medina-Selby, A., Masiarz, F. R., Hallewell, R. A., Chamberlain, S. H., Bradley, J. D., Lee, D., Steimer, K. S., Poulter, L., Burlingame, A. L., Esch, F., and Baird, A. (1988) *J. Biol. Chem.* **263**, 16471–16478
33. Laemmli, U. K. (1970) *Nature* **227**, 680–685
34. Tomasselli, A. G., Olsen, M. K., Hui, J., Staples, D. J., Sawyer, T. K., Heinrikson, R. L., and Tomich, C.-S. C. (1990) *Biochemistry* **29**, 264–269
35. Bathurst, I. C., Chester, N., Gibson, H. L., Dennis, A. F., Steimer, K. S., and Barr, P. J. (1989) *J. Virol.* **63**, 3176–3179
36. Steimer, K. S., Puma, J. P., Power, M. D., Powers, M. A., George-Nascimento, C., Stephans, J. C., Levy, J. A., Sánchez-Pescador, R., Luciw, P. A., Barr, P. J., and Hallewell, R. A. (1986) *Virology* **150**, 283–290
37. Leatherbarrow, R. J. (1987) *Elsevier Biosoft*, Cambridge, United Kingdom
38. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038
39. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., and Rosenberg, M. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 8903–8906
40. Hansen, J., Billich, S., Schulze, T., Sukrow, S., and Moelling, K. (1988) *EMBO J.* **7**, 1785–1791
41. Giam, C.-Z., and Boros, I. (1988) *J. Biol. Chem.* **263**, 14617–14620
42. Cousens, L. S., Shuster, J. R., Gallegos, C., Ku, L., Stempien, M. M., Urdea, M. S., Sánchez-Pescador, R., Taylor, A., and Tekamp-Olson, P. (1987) *Gene (Amst.)* **61**, 265–275
43. Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P., and Barr, P. J. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4642–4646
44. Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M., and Vogt, V. (1988) *J. Virol.* **62**, 1808–1809
45. Darke, P. L., Leu, C.-T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A. F., and Sigal, I. S. (1989) *J. Biol. Chem.* **264**, 2307–2312
46. Demsey, A., Collins, F., and Kawka, D. (1980) *J. Virol.* **36**, 872–877
47. Khan, A. S., and Stephenson, J. R. (1979) *J. Virol.* **29**, 649–656
48. Yoshinaka, Y., and Luftig, R. B. (1981) *Virology* **111**, 239–250
49. Tomasselli, A. G., Hui, J. O., Sawyer, T. K., Staples, D. J., Bannow, C., Reardon, I. M., Howe, W. J., DeCamp, D. L., Craik, C. S., and Heinrikson, R. L. (1990) *J. Biol. Chem.*, in press

The Journal of Biological Chemistry

PROTEINS: Structure, Function, and Genetics 6:324–337 (1989)

Recombinant HIV1 Protease Secreted by *Saccharomyces cerevisiae* Correctly Processes Myristylated *gag* Polyprotein

Sergio Pichuantes,[1] Lilia M. Babé,[1] Philip J. Barr,[2] and Charles S. Craik[1]
[1]*Departments of Pharmaceutical Chemistry and Biochemistry/Biophysics, University of California, San Francisco, California 94143 and* [2]*Chiron Corporation, 4560 Horton Street, Emeryville, California 94608*

ABSTRACT The protease of the human immunodeficiency virus type I (HIV1) was expressed both intracellularly and extracellularly in *Saccharomyces cerevisiae*. Intracellular expression of the protease was achieved by fusing a 179 amino acid precursor form of the protease to human superoxide dismutase (hSOD). Self-processing of the viral enzyme from the hybrid precursor was demonstrated to occur within the yeast host. Secretion of the protease was achieved by fusing the leader sequence of yeast α-factor to the precursor form of the protease or to the 99 amino acid mature form of the protease. Authentic and active forms of the retroviral enzyme were detected in yeast supernatants of cells expressing the precursor or the mature form of the protease. A D25E active site variant of the retroviral enzyme exhibited diminished autocatalytic activity when expressed intracellularly or secreted from yeast. The wild-type protease was active in an in vitro assay on the natural substrate, myristylated *gag* precursor, $Pr53^{gag}$. Correct processing of $Pr53^{gag}$ at the Tyr 138–Pro 139 junction was confirmed by amino terminal sequence analysis of the resulting capsid protein (CA, p24). The secreted protease was purified to homogeneity from yeast media using preparative isoelectric focusing and reverse-phase HPLC. Amino terminal sequence analysis showed a sequence beginning at amino acid 1 of the mature enzyme (Pro) and another sequence beginning at amino acid 6 (Trp). This shorter sequence may represent a natural autolytic product of the protease.

Key words: aspartyl protease, HIV/AIDS, yeast expression, polyprotein processing, α-factor, secretion

INTRODUCTION

Proteases are essential to the life cycle of all known retroviruses. The polyproteins produced by the compact viral genome must undergo limited proteolysis by processing enzymes, which release the various proteins required for the replication and assembly of the virus.[1] In order to ensure the correct processing of viral polyproteins, retroviruses frequently encode highly specific proteases[2] that are capable of recognizing unique amino acid sequences and hydrolyzing specific peptide bonds at the junctions between the viral proteins.[3] A potential means of intervention of the viral life cycle would involve inhibition of these hydrolytic events.

The etiological agent of AIDS is a retrovirus referred to as human immunodeficiency virus (HIV).[4] Genetic analysis of the molecular organization of the HIV genome has revealed that, as in other retroviruses, it consists of the *gag*, *pol*, and *env* major genes encoding several structural proteins and enzymatic activities.[5–7] These genes are expressed as polyproteins that require posttranslational proteolysis to yield the mature viral proteins. Processing of HIV type 1 (HIV1) *gag* and *gag/pol* precursors involves several proteolytic steps which are at least in part mediated by a viral protease encoded at the 5′ end of the *pol* gene[8] and expressed between *gag* and *pol* in the *gag/pol* precursor. Evidence that the viral proteolytic activity is crucial in viral replication is the loss of infectivity detected in virions encoding a mutant protease.[9] In order to gain a better understanding of the functional and structural aspects of this viral enzyme, reagent quantity levels of the authentic enzyme must be available for appropriate biophysical analysis. By using a heterologous expression system to overproduce the cloned gene product, the dependence upon natural sources of infected cells can be avoided. Furthermore, advantage can be taken of the expression system to produce genetically altered gene products that address specific structure–activity questions relating to the mode of action and substrate specificity of the enzyme.

Bacterial expression of various DNA fragments encoding different HIV1 protease precursors revealed that the 10 kDa retroviral enzyme is auto-

Received May 17, 1989; revision accepted July 28, 1989.
Address reprint requests to Charles S. Craik, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94608.

catalytically released from the higher molecular weight precursors.[10–12] This self-processing activity is abolished when the aspartic acid located at position 25 of the protease is replaced with an Ala,[13–15] a Thr,[16] a Tyr, or His[14] suggesting that the enzyme is a member of the aspartyl protease family. The recently reported three-dimensional structures of the HIV1 protease[17] and the related Rous sarcoma virus protease[18] confirms this classification. Cytosolic expression of the HIV1 *gag/pol* gene in yeast coupled with mutational analysis of the viral genome initially identified the DNA region that encodes the viral protease.[19] Despite this initial success, virtually all subsequent studies on the expression and mutagenesis of the protease have been performed in bacteria. Yeast expression of the HIV1 protease may provide an alternative method for overproducing and purifying this protein and its variants.

Saccharomyces cerevisiae has proven to be a useful host for the expression of recombinant proteins. Sophisticated plasmids have been constructed that include strong, regulated transcriptional control elements.[20–22] The availability of protease-deficient hosts and efficient transformation techniques[22–24] can be used to further increase the levels of production of foreign proteins. Sequences required for the secretion of the yeast pheromone α-factor have been identified and can be used for the secretion of various cloned gene products.[25–27] Heterologous expression of a number of human pathogen-associated polypeptides has been achieved using yeast hosts.[28–31] Furthermore, correct posttranslational modifications such as myristylation[32] and acetylation[33,34] of retroviral and human proteins have been shown to occur in yeast. These observations anticipate that this microorganism will continue to be a powerful tool for the biosynthesis and analysis of proteins that are associated with disease. To this end, we have expressed the HIV1 protease both intracellularly and extracellularly in *S. cerevesiae* and have established in vivo and in vitro assays using its naturally occurring substrates. This will aid in further characterization of the enzyme and permit the eventual design of inhibitors that may arrest the symptoms of AIDS.[35,36]

MATERIALS AND METHODS

Plasmid Constructions and Genetic Manipulations

Plasmid pPR99 encodes a polypeptide of 99 amino acids corresponding to the putative protease.[6] The DNA encoding this polypeptide was constructed from 13 oligonucleotides with 22 bp overlaps. Frequent codons of highly expressed yeast proteins were chosen whenever possible to optimize expression. Unique restriction sites were introduced throughout the coding sequence to facilitate future manipulations of the gene. Sequences were included at the 5′ end of the synthetic gene to regenerate an *Xba*I site and to encode the *KEX2* processing site (LeuAspLysArg) of α-factor. Sequences were included at the 3′ end to encode a stop codon (TAG) and to regenerate a *Sal*I cloning site (Fig. 1A).

Plasmid pPR179 encodes the 99 amino acids of the HIV1 protease together with 55 additional amino acid residues at the N-terminus and 25 additional amino acids at the C-terminus. The DNA encoding this 179 amino acid polypeptide was obtained by isolating the 522 bp *Bgl*II/*Bal*I DNA fragment from the cloned HIV1 proviral genome.[6,37] A synthetic adapter was added to the 5′ end of the DNA fragment to regenerate an *Xba*I restriction site, to encode the *KEX2* processing site and three additional residues of the *pol* sequence (PheArgGlu) and to regenerate a *Bgl*II restriction site. A synthetic adapter was added to the 3′ end of the DNA fragment to regenerate a *Bal*I restriction site, to encode an additional Pro residue of the *pol* gene and a stop codon (TAG), and to regenerate a *Sal*I restriction site (Fig. 1B). These reengineered *Xba*I/*Sal*I DNA fragments encoding the protease were used for all subsequent DNA constructions involving secretion of the protease from yeast.

For external localization of the protease, the two different *Xba*I/*Sal*I DNA fragments described above were ligated to the α-factor leader sequences and to the glucose regulated hybrid promoter ADH2/GAPDH.[33] The resulting DNA fragments were ligated into the yeast vector pBS24.1. This vector uses the α-factor terminator[25] to ensure transcriptional termination and the yeast genes, *leu2-d* and *ura3*, for selection. The defective *leu2-d* allelle[38] is used to increase the copy number of the plasmid. The 2μ sequences which include two inverted repeats are present for autonomous replication of the plasmid in yeast. The β-lactamase gene and ColE1 origin of replication are present for selection and autonomous replication in *Escherichia coli*. The plasmids were used to transform spheroplasts of the protease-deficient *S. cerevesiae* strain, AB110 (MATa, *leu2-3 112*, *ura3-52*, *pep4-3*, *his4-580* [*cir*°]). Leucine prototrophs were selected and grown in uracil-deficient media essentially as described.[33]

For internal localization of the protease, the plasmid pSOD/PR179 was made in which the *pol* sequences encoding the protease were fused to the C-terminus of human superoxide dismutase (hSOD). A synthetic adapter was designed for the 5′ end of the HIV1 DNA fragment to regenerate an *Nco*I restriction site, to encode a Met, Ala, and three amino acids of the *pol* gene product (PheArgGlu) and to regenerate a *Bgl*II restriction site. The *Bal*I/*Sal*I synthetic adapter described above was added to the 3′ end of the fragment encoding the *pol* sequences. This DNA construction was ligated to the *Nco*I/*Sal*I fragment of the vector pSODCF2[39] to generate pSODCF2/PR179. This recombinant plasmid creates an in-

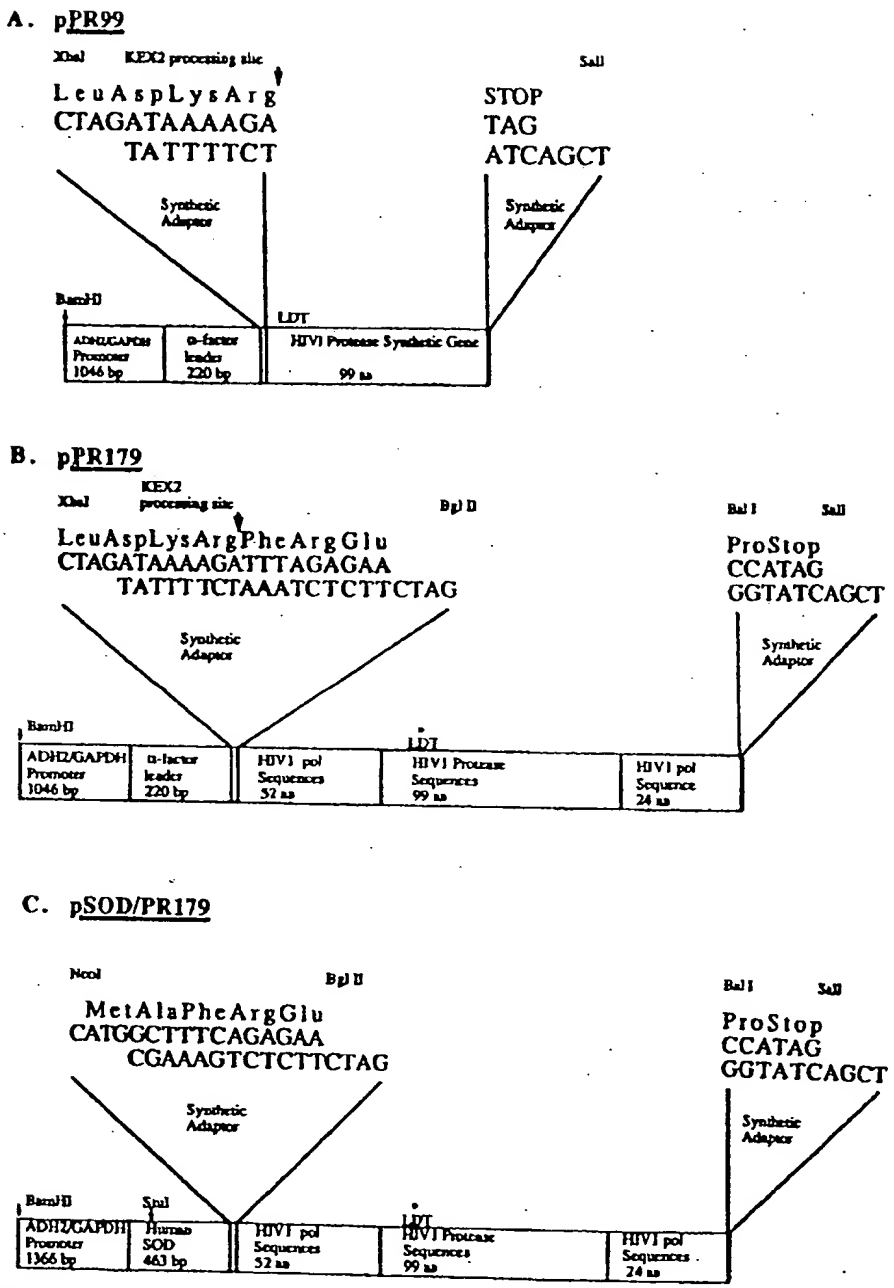


Fig. 1. Construction of yeast recombinant plasmids for the expression of HIV1 protease in *S. cerevisiae*. (A) Expression cassette for plasmid pPR99. The cassette encodes a 99 amino acid HIV1 protease. (B) Expression cassette for plasmid pPR179. The cassette encodes a viral protease containing 55 additional amino acid residues at its N-terminus and 25 additional amino acids at its C-terminus. Both plasmids include the yeast pheromone α-factor signal/leader sequence for secretion of the protease and the *KEX2* processing site which is denoted with an arrow. (C) Expression cassette for plasmid pSOD/PR179. The cassette encodes a fusion protein for internal expression in yeast. The hybrid protein consists of 333 amino acids with the C-terminus of hSOD fused to the N-terminus of the HIV1 viral protease precursor. These cassettes were ligated into the yeast expression vector pBS24.1 which included β-lactamase, *leu2-d* and *ura3* markers, 2μ sequences, and the α-factor transcription terminator. LDT represents the amino acids Leu, Asp, and Thr located at the putative active site of the HIV1 protease. Arterisks indicate sites where aspartic acid 25 was replaced with a glutamic acid. All recombinant plasmids include the ADH2/GAPDH promoter to regulate the expression of the viral protease.

frame fusion of hSOD to the *pol* gene products with the carboxyl terminal Asn of hSOD being replaced by a Met and Ala. The 833 bp *StuI–SalI* fragment of pSODCF2/PR179 was isolated and ligated into the plasmid pSI1[22] to provide the ADH2/GAPDH promoter. Finally, the *BamHI/SalI* fragment of the resulting plasmid pSI1/PR179 was introduced into the yeast vector pBS24.1 described above. The resulting recombinant plasmid, pSOD/PR179 (Fig. 1C), was prepared and used to transform spheroplasts of the protease-deficient strain *S. cerevisiae* AB116 (*prc1-407, prb1122, pep4-3, leu2, trp1, ura3-52,* [*cir*°]).[33] Yeast transformation, selection, and growth of recombinants were achieved as described elsewhere[20] except that YEP media[40] supplemented with 2% glucose was used instead of uracil-deficient media for intracellular expression from the ADH2/GAPDH promoter.

Oligonucleotide Synthesis and Site-Directed Mutagenesis

Oligonucleotides were synthesized using solid-phase phosphoramidite chemistry on an Applied Biosystems 380A DNA synthesizer. The deprotected oligonucleotides were purified by electrophoresis on 15% polyacrylamide gel containing 8 M urea and eluted overnight in 1 ml of 0.1 M Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM EDTA. Samples were chromatographed through a Sep-Pak C_{18} cartridge, lyophilized, and suspended in water. The mutagenic oligonucleotide 5′ AGCTCTATTAGAAACAGGAGCAGAT 3′ was used to prime the synthesis of a DNA strand encoding an HIV1 protease template with a glutamic acid codon (GAA) in place of the aspartic acid codon (GAT) at position 25 in the viral protease. Mutagenesis, transformation of *E. coli* CJ236, screening, and template sequencing were done as previously reported.[41] Yeast plasmids encoding mutant protease were obtained after replacing the 522 bp *BglII/SalI* fragment of pPR179 and pSOD/PR179 with identical DNA fragments except for the glutamic acid codon replacement. The resulting engineered plasmids were denoted pPR179 D25E and pSOD/PR179 D25E, respectively, according to standard nomenclature for mutant proteins.[42]

Detection of HIV1 Protease Expressed in Yeast

Cultures of *S. cerevesiae* AB110 grown in uracil deficient media for 24–48 hours were centrifuged at 7000*g* for 15 minutes and proteins in 10 ml of supernatant were precipitated with 5 ml of 50% trichloroacetic acid which contained 2 mg/ml deoxycholate. After incubation at 4°C for 30 minutes, sediments were collected by centrifugation at 12,000*g* for 30 minutes, washed with acetone, dried, and suspended in 40 µl of sample buffer.[43] Yeast AB116 cells were harvested as described for AB110 and 1 OD_{650} of cells were lysed by repeated boiling and freezing in 30 µl of 63 mM Tris-HCl, pH 6.8, 50 mM DTT, 10% glycerol, 3% SDS, 0.02% bromophenol blue. Proteins were fractionated by SDS–PAGE in 15% polyacrylamide gels and the proteins transferred electrophoretically to nitrocellulose filters for 1 hour at 75 V in a Bio-Rad Trans-Blot apparatus. The HIV1 protease was detected on immunoblots using rabbit polyclonal antibodies raised against concentrated yeast supernatant from cells expressing an HIV1 polypeptide encompassing the 78 C-terminal amino acids of the protease and the 37 N-terminal residues of the reverse transcriptase.[20] A goat anti-rabbit antibody conjugated to horseradish peroxidase (HRP, Boeringer/Mannheim) was used for immunodetection. The rabbit polyclonal antibodies recognize secreted yeast proteins other than the viral protease since the immunogen was only partially purified prior to immunization. The human SOD/PR fusion polypeptide was detected using mouse monoclonal antibodies raised against hSOD as primary antibodies and goat anti-mouse antibodies conjugated to HRP (Tago) as secondary antibodies.

In Vitro Assay for HIV1 Protease

HIV1 protease activity was assayed by incubating yeast recombinant myristylated $Pr53^{gag}$ with the protease and monitoring the formation of matrix protein (MA p17) and capsid protein (CA p25/p24)[44] using SDS–PAGE and immunoblotting. The specific viral proteins were visualized on immunoblots using serum from AIDS infected patients as the primary antibody and goat anti-human antibodies conjugated to HRP (Tago) as the second antibody. Total lysates of HIV1-infected cells[45] were included as standards to visualize HIV1 viral proteins. Reactions were carried out in 10 mM Tris-HCl, pH 7.0, 130 mM NaCl, 1 mM EDTA, and 1 mM phenylmethylsulfonylfluoride (PMSF) at 20°C using 1 µg of $Pr53^{gag}$ as substrate.[32] Samples of $Pr53^{gag}$ were incubated with HIV1 protease in concentrated media from transformed yeast AB110 or in partially purified lysates from transformed yeast AB116.

Extracellularly expressed protease was concentrated from the media of yeast AB110 expressing either pPR99 or pPR179. Yeast cultures were grown to saturation for 24 or 48 hours prior to harvesting the cells. The cells were removed from the media at 7000*g* for 15 minutes, and the supernatant then concentrated at 4°C approximately 20-fold using either a Centricon 10 or a Centriprep 10 (Amicon) with a molecular weight cutoff of 10,000.

Partial purification of the internally expressed protease was achieved from transformed yeast AB116 cells as follows. Cell cultures were pelleted by centrifugation and frozen at −20°C. The cells were resuspended on ice in 50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 1 mM DTT, 1 mM PMSF, 100 mM KCl, and 0.5% Triton X-100 and sonicated using 30

second bursts until the viscosity of the solution that resulted from liberation of chromosomal DNA was significantly reduced. The lysed cells were then clarified by centrifugation at 5000*g* for 20 minutes. To remove DNA from the sample, neutralized protamine sulfate was added to 0.5% (w/v), the solution incubated at room temperature for 15 minutes, and centrifuged at 7000*g* for 15 minutes. The supernatant was then adjusted to 15% ammonium sulfate using a standard solution. After 30 minutes at room temperature, the precipitate was removed by centrifugation and the supernatant was brought to 40% ammonium sulfate to precipitate the protease. Following a 1 hour incubation at room temperature, and centrifugation at 7000*g* for 15 minutes, the pellet was collected, resuspended in 10 mM Tris-HCl, pH 7.5, and dialyzed exhaustively against 10 mM Tris-HCl, pH 7.5 buffer. This resulted in a 20-fold purification of the protease as monitored by OD_{280}.

Amino-Terminal Amino Acid Determinations

Sequential Edman degradation of the viral products resulting from specific hydrolysis by the viral protease was performed using an Applied Biosystems Model 470A gas-phase protein sequenator. The sequenator was equipped with an on-line phenylthiohydantoin (PTH) amino acid analyzer. The PTH-amino acids were separated by reverse-phase chromatography on a Brownlee C_{18} column.[46] To determine the amino-terminal sequence of the processed capsid protein p24, 15 μg of myristylated $Pr53^{gag}$ was digested with concentrated media from yeast transformed with either pPR99 or pPR179 using the conditions described for the in vitro assay. The length of incubation ranged from 10 to 48 hours. The products of the digestion were then separated by SDS–PAGE. After electrophoresis, proteins were electroblotted onto PVDF membranes (Immobilon-P Transfer Millipore) in 10 mM 3-[cyclohexylamino]-1-propanesulfonic acid (CAPS), pH 11.0, 10% methanol. Membranes were washed in deionized water, stained with 0.1% Coomassie Brilliant Blue R-250 in 50% methanol, destained, dried, and stored at −20°C. Strips of PVDF membrane containing the capsid protein were cut out and the protein was sequenced following standard procedures.[47] Amino-terminal sequence analysis of the HIV1 protease was carried out on purified PR99.

Purification of HIV1 Protease Secreted Into Yeast Media

Twelve liters of yeast media from cells transformed with pPR99 that had been grown for 48 hours was concentrated to 2 liters using a Pellicon tangential flow concentrator (Amicon) equipped with a PTGC cassette with a 10,000 molecular weight cutoff. The sample was concentrated further to 200 ml on an Amicon ultrafiltration cell using a YM5 membrane with a 5000 molecular weight cutoff. Following exhaustive dialysis against water, the solution was adjusted to 0.1% Triton X-100 and 1 mM PMSF. Protein separation by preparative isoelectric focusing was accomplished on a Rotofor unit (Bio-Rad). A 55 ml sample of concentrated and dialyzed media was mixed with 1.5 ml of pH 3–10 ampholytes (40% solution, Bio-Rad), and separated on the Rotofor according to the manufacturer's indications. Twenty fractions of approximately 2 ml each were collected and their pH and OD_{280} were measured. The presence of the viral protease in these fractions was determined on immunoblots and found to be localized in fractions which range from pH 9 to 12. These fractions were pooled, placed in dialysis bags, and concentrated against polyethylene glycol (PEG). The PEG was then removed by dialysis against water. The concentrated sample was then subjected to reverse-phase HPLC using a C_3 column (Protein Plus, Dupont) developed with a water:acetonitrile gradient containing 0.1% trifluoroacetic acid (TFA). The HIV1 protease eluted at approximately 55% acetonitrile. Fractions of 1 ml each were collected and those containing the protease, as monitored by SDS–PAGE and silver staining, were collected and lyophilized. The approximate yield of highly purified material from 12 liters of media was 1 mg.

The specific activity of the purified protease was measured using two synthetic peptides and an HPLC-based discontinuous assay.[48] The sequences of the peptides were Ala Thr Leu Asn Phe Pro Ile Ser Pro Trp and Arg Ser Leu Asn Tyr Pro Gln Ser Lys Trp. Aliquots of peptide substrates (10 μg) were incubated with purified protease (0.5 μg) at 37°C for 1 minute to 24 hours in 50 mM sodium phosphate buffer pH 5.5 containing 4 mM EDTA, 20 mM DTT, and 25 mM NaCl in a final volume of 20 μl. The reaction mixtures were fractionated by reverse-phase HPLC on a Vydac C_{18} column (250 × 4.6 mm) using an aqueous gradient of acetonitrile from 0 to 100% containing 0.1% TFA and a flow rate of 1 ml/minute. The peptides were detected by UV absorbance at 215 mm. The amount of peptide substrate was calculated from the absorption profiles and plotted versus time. A specific activity was calculated from the slope of the curve in the linear range.

RESULTS

The expression and secretion of HIV1 protease in yeast were accomplished by cloning two different DNA fragments encoding the viral enzyme into a yeast vector that provided a strong inducible promoter for expression and the signal peptide and processing signals of the yeast pheromone α-factor for secretion. Recombinant plasmids pPR99 and pPR179 encode the 99 amino acid protease and a 179 amino acid protease precursor, respectively (Fig. 1A and B). A 10 kDa polypeptide was detected immunologically in yeast media from cells expressing ei-

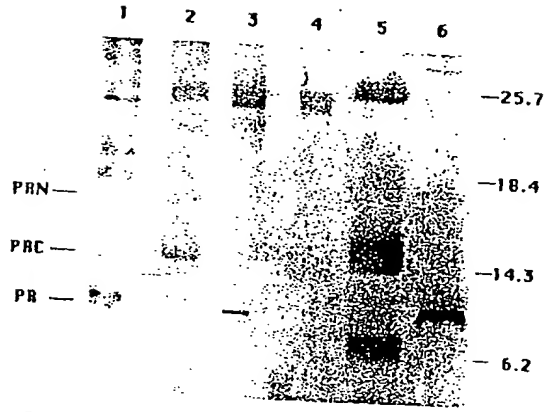


Fig. 2. Immunoblot analysis of HIV1 protease expressed in *S. cerevisiae* AB110. Yeast cells harboring recombinant plasmids pPR99 (lane 1), pPR179 D25E (lane 2), and pPR179 (lane 3) or the yeast vector pBS24.1 (lane 4) were grown in uracil-deficient media at 30°C. After 48 hours of incubation, cultures were centrifuged and proteins in yeast supernatants were precipitated with TCA and electrophoresed in a 15% polyacrylamide gel containing SDS. Immunoblot analysis was performed using a rabbit polyclonal serum raised against an HIV1 polypeptide encompassing the 78 C-terminal amino acids of the protease and the 37 N-terminal residues of the reverse transcriptase.[20] A bacterial expressed HIV1 protease[48] was included in lane 6. Prestained molecular weight markers were loaded in lane 5. PR is the protease, PRN is the protease plus 55 amino acid N-terminal extension, and PRC is the protease plus 25 amino acid C-terminal extension. The level of total secreted protein varied for each of the different transformed yeast shown in lanes 1 through 4 as determined from Coomassie Blue-stained SDS–PAGE of the samples. The excess protein loaded in lane 1 relative to lanes 2 and 3 presumably accounts for the background band that nearly comigrates with PRC in lane 1.

ther pPR99 (Fig. 2, lane 1) or pPR179 (Fig. 2, lane 3). This protein comigrates with HIV1 protease expressed in bacteria[48] as shown in lane 6. A 10 kDa band is not observed in media from yeast transformed with a plasmid that does not encode the protease (Fig. 2, lane 4). The predicted molecular weight of the 99 amino acid polypeptide is 10,792, which is in agreement with the observed size as determined by its migration in SDS polyacrylamide gels relative to proteins with known molecular weights. This shows that the protease is expressed in yeast from the plasmid construction and indicates that the α-factor leader sequence provides an efficient system to export the viral enzyme from the cell. The use of α-factor sequences to export recombinant proteins from yeast has been reported for other gene products including human epidermal growth factor,[24] carboxypeptidase A,[26] and human fibroblast growth factor.[3] However, the HIV1 protease differs from these proteins in that it is normally exported from cells by a mechanism involving viral budding.

The plasmid pPR179 encodes an HIV1 protease containing 55 additional amino acid residues at its N-terminus and 25 additional amino acids at its C-terminus. However, a protein band of approximately 10 kDa was detected in media of yeast transformed with this plasmid (Fig. 2, lane 3). This suggests that the protease precursor undergoes proteolytic processing resulting in the 10 kDa mature protease. This event probably represents an autocatalytic process mediated by the protease itself since no mature protease was detected in yeast supernatants of cells harboring a plasmid encoding a variant form of the protease pPR179 D25E (Fig. 2, lane 2). In this variant enzyme, the active site aspartic acid has been replaced with a glutamic acid in order to reduce the activity of the protease with minimal perturbation of the enzyme structure. The autoprocessing event that is occurring either along the secretory pathway of the yeast or in the extracellular media has also been observed intracellularly in bacteria.[10,11,14,15,48–50]

In the yeast supernatant of cells expressing the variant protease a 13.5 kDa protein band was revealed by the rabbit immune sera. This polypeptide, which we refer to as PRC, may represent a 124 amino acid residue intermediate of 13,534 that would be formed if the 25 amino acid C-terminal extension was not removed while the N-terminal 55 amino acid extension was removed. A minor band of approximately 16.5 kDa was also detected, PRN, which would result if the C-terminal extension was removed but the N-terminal extension was not removed leading to the 154 amino acid intermediate of 16,950 Da. The full-length precursor of 19.7 kDa is not observed, presumably due to its extreme lability. Yeast expression of pPR179 D25E results in the presence of the two intermediates PRC and PRN suggesting that the Asp to Glu substitution leads to a protease with reduced, but measurable self-processing activity.

When a DNA fragment encoding the 179 amino acid precursor was cloned in a yeast vector for internal localization of the protease, the 10 kDa protease band was not detected immunologically (results not shown). Since no difference in cell growth resulted upon transformation of yeast with the pPR179 expression vector, the viral protease precursor presumably was degraded by intracellular proteases. Expression of the HIV1 protease in yeast required fusion of the precursor sequences to the DNA sequences encoding hSOD. Analysis of viral protein products expressed in *S. cerevisiae* strain AB116 harboring either pSOD/PR179 or pSOD/PR179 D25E plasmids is shown in Figure 3. These plasmids encode a fusion protein of 333 residues with a calculated molecular weight of 35,555, composed of 153 amino acids of hSOD with the C-terminal Asn replaced with a Met, one spacer amino acid (Ala), and 179 amino acids of the protease precursor. A 10 kDa polypeptide was detected in lysates of yeast cells containing the pSOD/PR179 recombinant plasmid (Fig. 3A, lane 4). The 10 kDa protein comigrates with the HIV1 protease expressed in bacteria (Fig.

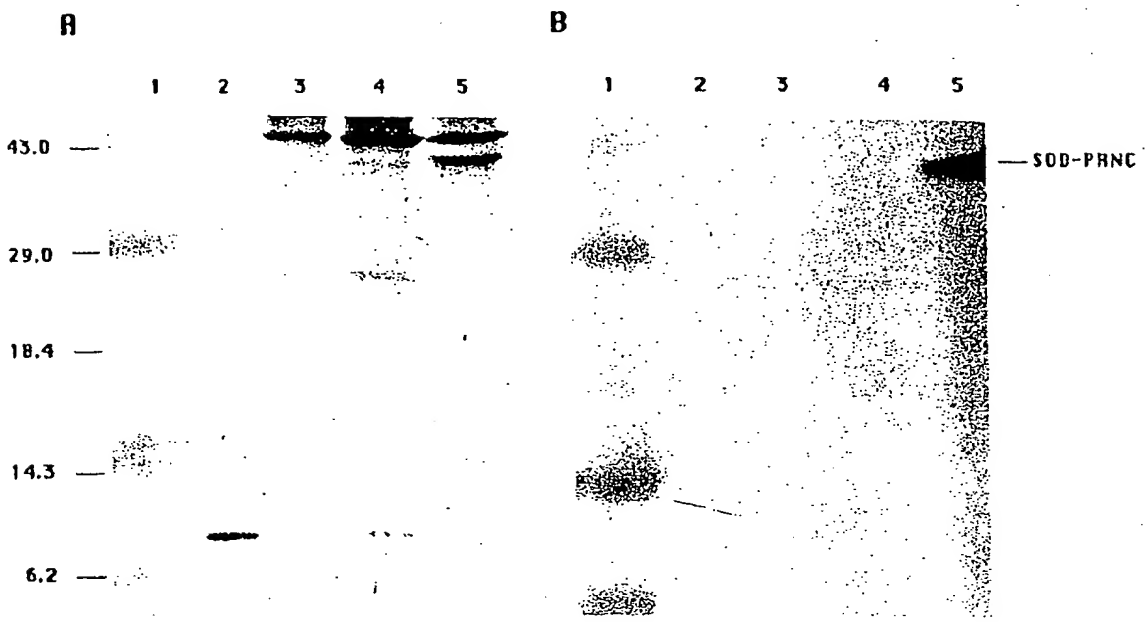


Fig. 3. Detection of HIV1 protease expressed in *S. cerevisiae* AB116 by immunoblot of yeast total cell lysates. (A) Immunoreactivity with rabbit serum containing antibodies against the HIV1 protease. (B) Immunoreactivity with mouse monoclonal antibodies to human superoxide dismutase. Cells grown at 30°C in YEP media supplemented with 2% glucose for 48 hours were lysed and processed as described in Materials and Methods. Lane 1: Pre-stained molecular weight markers. Lane 2: Bacterial expressed HIV1 protease. Lane 3: Lysate of yeast containing pBS24.1. Lane 4: Lysate of yeast expressing pSOD/PR179. Lane 5: Lysate of yeast expressing pSOD/PR179 D25E. SOD-PRNC is the 333 amino acids hybrid precursor consisting of hSOD and HIV1 *pol* amino acid sequences.

3A, lane 2). This suggests that the viral proteolytic processing occurred within the yeast host to generate a protease of correct molecular weight. The 10 kDa band is not observed in lysates of yeast cells transformed with a plasmid that does not encode the protease (Fig. 3A, lane 3). Lysates of yeast cells harboring pSOD/PR179 D25E do not show the 10 kDa protein band (Fig. 3A, lane 5) although a 35 kDa band (SOD-PRNC) is detected by antibodies to the HIV1 protease (Fig. 3A, lane 5) and by antibodies to hSOD (Fig. 3B, lane 5). This band probably represents a fused hSOD-HIV1 protease precursor that was not processed by the variant protease. No such precursor was detected in cells harboring pSOD/PR179 (Fig. 3A, lane 4; Fig. 3B, lane 4) suggesting that this molecule is rapidly cleaved within the yeast cell to form the mature protease. A 209 amino acid intermediate of molecular weight 22,021 representing hSOD bound covalently to HIV1 sequences located immediately upstream of the amino terminus of the protease is not observed. Presumably this hybrid molecule is very susceptible to the action of yeast proteases and is rapidly degraded.

The yeast expressed HIV1 protease was shown to be enzymatically active in an in vitro assay that uses recombinant myristylated $Pr53^{gag}$ precursor as a substrate.[32] Myristylation is known to occur in a number of mammalian retroviral *gag* precursors[51] and may be essential for viral polyprotein transport to the plasma membrane and subsequent virion formation.[52,53] Figure 4A shows that an active HIV1 protease is present in supernatants of yeast cells expressing either pPR99 (Fig. 4A, lanes 5,6) or pPR179 (Fig. 4A, lanes 7,8), as evidenced by the generation of major capsid (p24) and matrix protein (p17) that comigrate with their counterparts expressed in vivo (lane 1). At least two other minor reactive species of approximately 25 and 39 kDa, that may represent processing intermediates,[19,54] were also detected by the AIDS sera. Neither $p16^{gag}$ nor its putative processed products, nucleocapsid protein (NC, p9) and $p7^{gag}$, were detected, presumably due to their rapid turnover or low immunogenicity. The highly purified $Pr53^{gag}$ (Fig. 4A, lanes 3,4) was nonspecifically cleaved by a yeast protease present in the media of yeast transformed with a plasmid not encoding the viral protease (Fig. 4A, lanes 9,10). The HIV1 protease is insensitive to 1 mM PMSF, which was included in the in vitro assays to avoid nonspecific degradation, and is partially or totally inhibited by 1 or 10 mM pepstatin A, respectively (results not shown).

When the same protocol is followed to test the enzymatic activity of the internally produced HIV1 protease (Fig. 4B), essentially identical results are obtained. Capsid and matrix proteins are formed

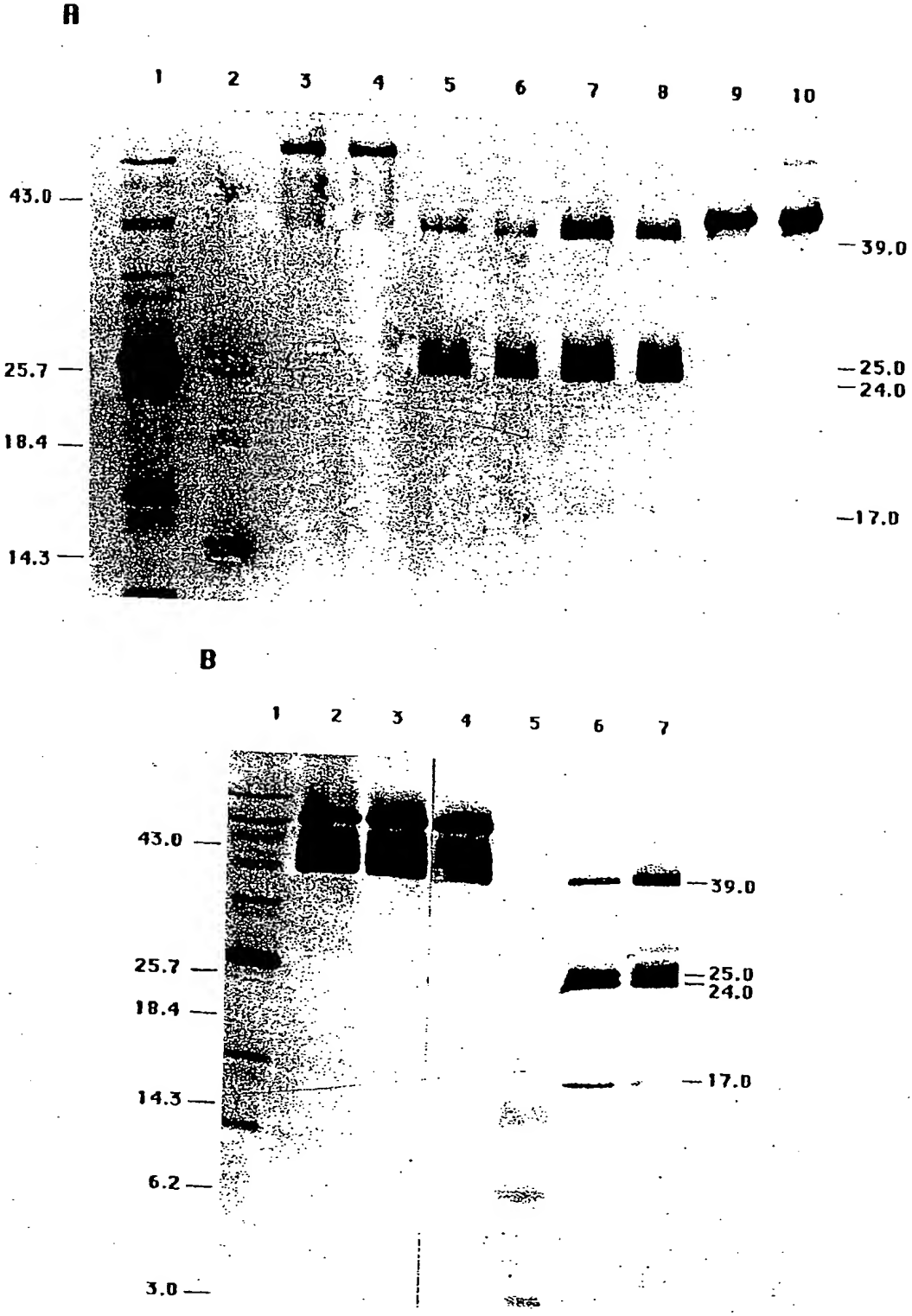


Fig. 4. Legend appears on page 332.

TABLE I. Processing Sites of HIV1 *gag* and *gag/pol* Polyproteins*

Polypeptide junction	Cleavage site								
	P4	P3	P2	P1		P1′	P2′	P3′	P4′
	135								142
Matrix protein p17 (MA) ↓ capsid protein p25/p24 (CA)	S	Q	N	Y	↓	P	I	V	Q
	366								373
Capsid protein p25 (internal cleavage)†	A	R	V	L	↓	A	E	A	M
	380								387
Capsid protein p25 (CA) ↓ p16	A	N	I	M	↓	M	Q	R	G
	451								458
p16: nucleocapsid protein p9 (NC) ↓ p7	P	G	N	F	↓	L	Q	S	R
	53								60
Pol precursor ↓ protease (PR)	S	F	N	F	↓	P	Q	I	T
	58								65
Protease (PR) ↓ protease p10* (internal cleavage)‡	Q	I	T	L	↓	W	Q	R	P
	152								159
Protease (PR) ↓ reverse transcriptase p66 (RT)	T	L	N	F	↓	P	I	S	P
Reverse transcriptase p66 (RT): p51 ↓ p15	No information available								
	712								719
Reverse transcriptase p66 (RT) ↓ integration protein p32 (IN)	R	K	V	L	↓	F	L	N	G

*Amino acid sequence data are from reference 6. The number above the one-letter code amino acids represent the positions within the *gag* or *pol* precursors. Cleavage sites were defined on the basis of amino-terminal amino acid determinations obtained in this work as well as on protein sequence information reported previously.[6,10,54,58-61]

†The internal cleavage to generate p24 from p25 occurs at the C terminus of p25.

‡Whether this cleavage is mediated by the viral encoded protease and/or has some biological significance is currently investigated in our laboratory. The nomenclature or amino acid residues at the cleavage sites has been reported elsewhere.[6] The standardized two-letter notation proposed for retroviral proteins[44] was followed when possible.

along the course of the incubation when extracts were used from yeast cells transformed with pSOD/PR179 (Fig. 4B, lanes 6,7). Long-term incubations resulted in the disappearance of the 39 and 25 kDa polypeptides to form the 24 kDa capsid protein and the 17 kDa matrix protein (Fig. 4B). These results not only illustrate the nature of intermediates of those products, but also provide evidence that the viral protease can cleave capsid protein p25 at its leucine–alanine junction.[6] Cleavage of this peptide bond by a synthetic HIV1 protease has been demonstrated.[55] Moreover, removal of the 14 carboxyl residues from capsid protein p25 to yield capsid protein p24 has also been reported to occur within HIV1-infected cells.[54] As expected, no specific processing of $Pr53^{gag}$ was detected when extracts were used that contained a plasmid that did not express the viral protease (Fig. 4B, lanes 2,3). When $Pr53^{gag}$ was incubated with extracts from cells expressing the variant protease, PR D25E, no processing activity was detected (results not shown), suggesting that aspartic acid 25 is required for the enzymatic activity of the protease on $Pr53^{gag}$ precursor in the in vitro assay.

To unequivocally demonstrate that the yeast expressed HIV1 protease can correctly process myristylated $Pr53^{gag}$, the N-terminal sequence of the enzymatically produced capsid protein p24 was determined. The sequence Pro-Ile-Val-Gln-Asn-Leu-Gln-Met-Val was obtained and compared with the predicted amino acid sequence of the capsid protein p24. Both polypeptides show identical residues at all positions where the amino acid sequence was determined for capsid protein p24, confirming that the viral enzyme expressed and secreted in yeast efficiently recognized and cleaved the predicted tyrosine–proline peptide bond between the matrix protein and capsid protein (Table I).

The scheme chosen for purification of the viral protease takes advantage of two unusual properties

Fig. 4. Immunological analysis of in vitro processing of the HIV1 $Pr53^{gag}$ precursor by yeast expressed HIV1 protease. (A) Secreted protease. One microgram of myristylated $Pr53^{gag}$ precursor[32] was incubated at 20°C with yeast concentrated supernatants in 130 mM NaCl, 1 mM EDTA, 1 mM PMSF, 10 mM Tris-HCl, pH 7.0–7.5. After 48 hours of incubation, samples were taken and processed for immunoblotting using AIDS patient sera. HIV1 infected total cell lysate[45] (lane 1). Molecular weight markers (lane 2). $Pr53^{gag}$ incubated in buffer pH 7.0 (lane 3) or pH 7.5 (lane 4). $Pr53^{gag}$ incubated at pH 7.0 (lane 5) or pH 7.5 (lane 6) with yeast supernatants of cells expressing pPR99. $Pr53^{gag}$ incubated at pH 7.0 (lane 7) or pH 7.5 (lane 8) with yeast supernatants of cells expressing pPR179. $Pr53^{gag}$ incubated at pH 7.5 (lane 9) or pH 7.0 (lane 10) with yeast supernatant of cells harboring the vector pBS24.1. (B) Intracellurly expressed protease Assays were carried out in 130 mM NaCl, 1 mM EDTA, 1 mM PMSF, 10 mM Tris-HCl, pH 7.0, at 20°C for 24 or 48 hours. Products of digestion were fractionated on 15% polyacrylamide gels containing SDS. Immunoblot analysis were performed using AIDS patient sera. HIV1 infected cell total lysate (lane 1). $Pr53^{gag}$ incubated for 24 hours (lane 2) or 48 hours (lane 3) with cell extracts of yeast containing the pBS24.1 vector. $Pr53^{gag}$ incubated in buffer for 48 hours (lane 4). Molecular weight markers (lane 5). $Pr53^{gag}$ incubated for 24 hours (lane 7) or 48 hours (lane 6) with extracts of cells expressing pSOD/PR179.

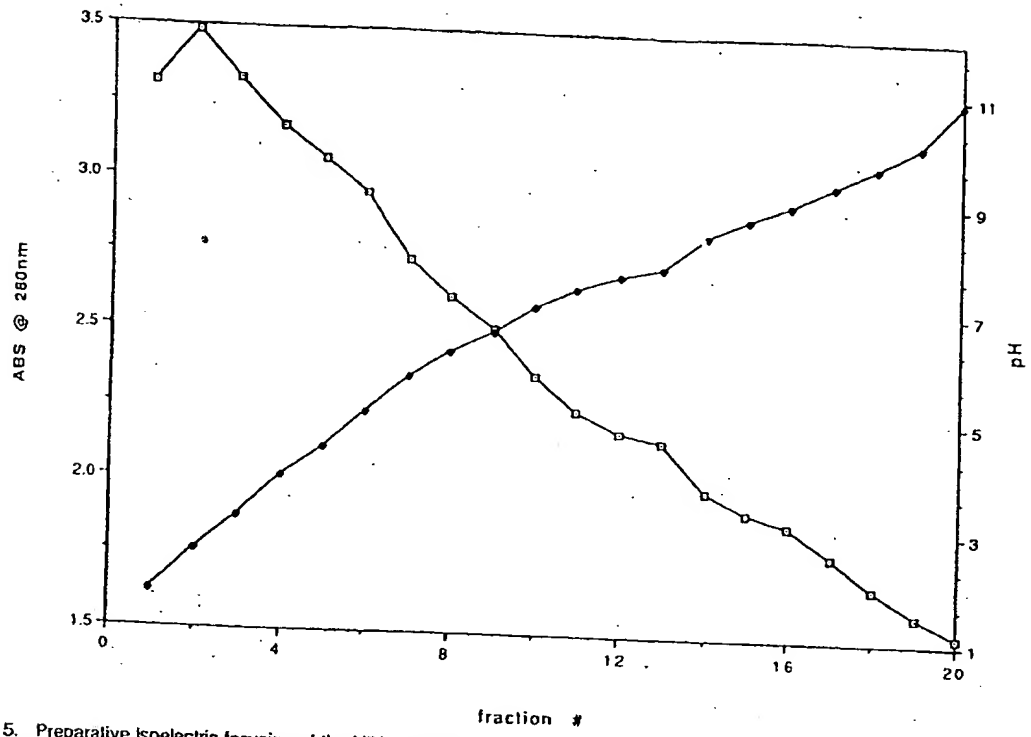


Fig. 5. Preparative isoelectric focusing of the HIV1 protease. Concentrated yeast culture supernatant of cells expressing pPR99 was dialyzed exhaustively against water, adjusted to 0.1% Triton X-100, 1 mM PMSF, and 1% ampholytes (pH 3–10 range), and loaded on a Rotofor unit (Bio-Rad). The pH (closed diamonds) and the OD_{280} (boxed dots) were determined for each of the 20 fractions obtained.

of this protein, namely the high isoelectric point and the very hydrophobic nature. Consistent with a predicted isoelectric point of 9.95, the protease migrated to a region between pH 9 and 12 in a preparative isoelectric focusing separation (Fig. 5). Few contaminating proteins were found in this pH range (fractions 17 to 20) since the majority of yeast proteins have much lower isoelectric points. This apparatus permits us to collect the protease in a small volume which is further concentrated against PEG, dialyzed, and subjected to reverse-phase HPLC. The C_3 column used was developed using a water:acetonitrile gradient, and, as expected for a hydrophobic protein, the protease eluted at an acetonitrile concentration of approximately 55% (Fig. 6A). Fractions eluting between 26 and 29 minutes consisted of homogeneous HIV1 protease as judged by silver staining of an aliquot separated on SDS–PAGE (Fig. 6B, lane 7). Incubation of the purified protease with the synthetic peptide Ala Thr Leu Asn Phe Pro Ile Ser Pro Trp at pH 5.5 results in the specific hydrolysis of 20 nmol of peptide/minute/mg protease. Incubation of the peptide Arg Ser Leu Asn Tyr Pro Gln Ser Lys Trp under the same conditions results in the specific hydrolysis of 54 nmol/minute/mg protease. These values are in the range of those observed for enzyme preparations of bacterially expressed HIV protease on similar substrates.[48,61]

To determine whether the yeast-secreted HIV1 protease represents the 99 amino acid residue mature product, amino-terminal sequence analysis was performed on the purified material. Shown in Figure 7 are the amino-terminal sequences that were determined for the yeast expressed HIV1 protease. Identical amino acids to those of the HIV1 protease[6] were found for the 10 amino-terminal residues of the yeast expressed protease. A polypeptide sequence starting at tryptophan 6 was also determined by sequence analysis (Fig. 7). This 94 amino acid polypeptide, p10*(94), which lacks the 5 amino-terminal residues, may have been generated autocatalytically and may represent a biologically significant viral product.

DISCUSSION

Heterologous expression of a virally encoded protein from a portion of the viral genome can provide reagent levels of the protein for biochemical analyses. Besides avoiding the biohazard of working with large quantities of virus-infected cells, the expression system permits an analysis of the protein using current methods of molecular genetics. Bacterial ex-

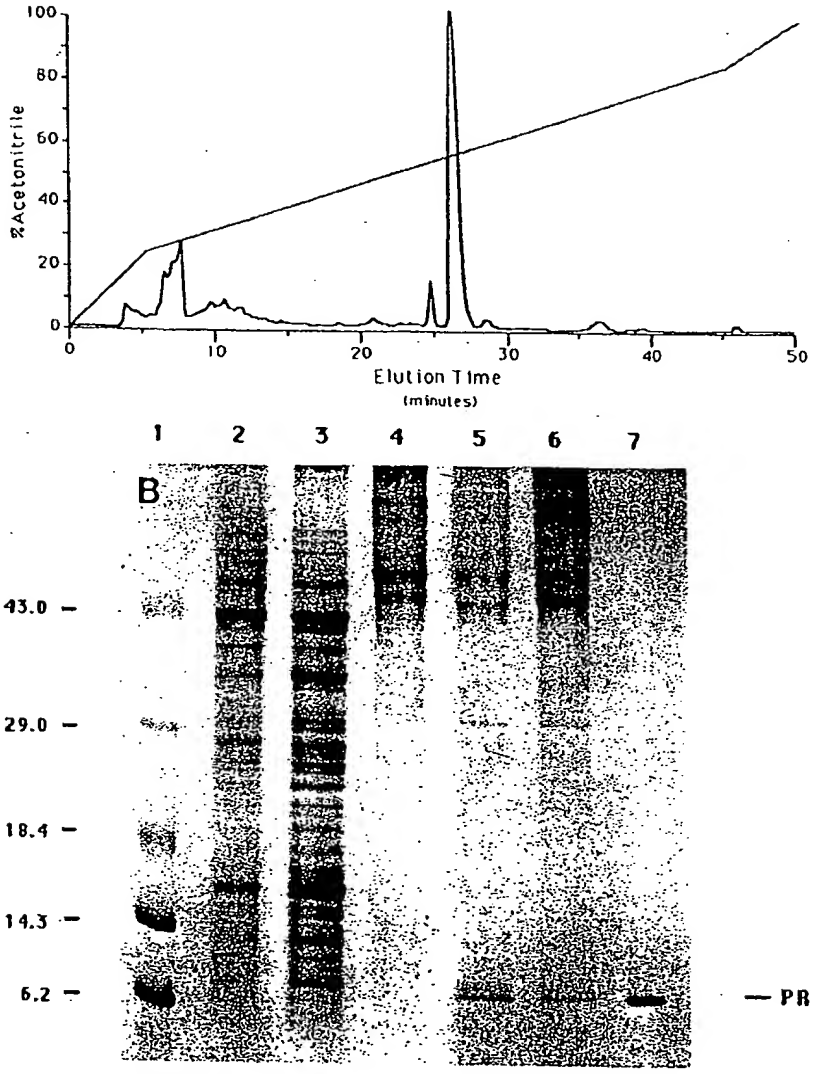


Fig. 6. Reverse-phase HPLC of the HIV1 PR99 protease. 6A. Samples of protease-containing fractions, isolated by isoelectric focusing, were loaded on a C_3 column. Material was eluted using a water:acetonitrile gradient (in 0.1% TFA) of 0–25% for 10 minutes, 25–85% for 40 minutes, and 85–100% for 5 minutes at a flow rate of 1 ml/min. Elution was monitored at 280 nm. 6B. Coomassie Blue-stained 15% polyacrylamide gel containing SDS of yeast expression products. Moleculars markers (lane 1). Yeast lysate of strain AB116 harboring the vector pBS 24.1 with no protease insert (lane 2). Yeast lysate of strain AB116 harboring the plasmid pSOD/179 (lane 3). Concentrated yeast media of strain AB110 harboring pBS 24.1 with no protease insert (lane 4); pPR99 (lane 5); pPR179 (lane 6). Purified HIV1 protease, PR99, 0.4 μg (lane 7).

pression systems have been established for biosynthetic production of the HIV1 protease.[10–16,48] However, internal localization of the protease has resulted in toxic side effects of the protease, instability of the enzyme to bacterial proteases, solubility problems (inclusion bodies), and difficulties associated with purifying low levels of a protein from a total *E. coli* extract. These difficulties have frustrated attempts to isolate reagent quantities of the HIV1 protease and its variants. Attempts to secrete the protease from *E. coli* which may circumvent some of these problems have thus far met with little success. We therefore took advantage of the eukaryotic microorganism *Saccharomyces cerevisiae* to express the retroviral aspartyl protease. *S. cerevisiae* has been used to efficiently express a number of engineered proteins intracellularly[19,20,22,24,30,32] or extracellularly.[21,25,27,33] When a DNA fragment encoding this HIV1 protease was fused in-frame to hSOD for internal localization in yeast, a mature

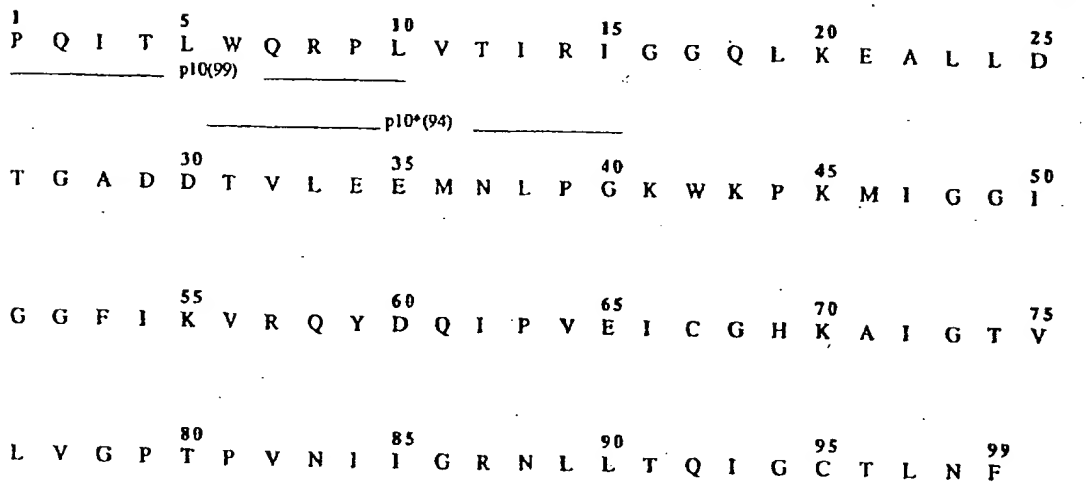


Fig. 7. Amino-terminal sequence analysis for purified HIV1 protease. Sequencing was performed on an Applied Biosystems Model 470A gas-phase protein sequenator equipped with an on-line phenylthiohydantoin (PTH) amino acid analyzer. PTH-amino acid separations were performed by reverse-phase chromatography on a Brownlee C_{18} column. The predicted amino acid sequence of the HIV1 protease[8] is written using the one-letter amino acid code. The amino acids determined by the sequence analysis are underlined and agree with the predicted amino acid sequence of the *pol* encoded protease. The name and predicted length of each polypeptide are indicated in the center of the underline. p10(99) is the mature HIV1 protease. p10*(94) is the protease lacking the first 5 amino-terminal residues.

form of the viral enzyme was detected immunologically. This result indicates that the autocatalytic event that has been shown to occur in bacteria[10–12,48] can also take place within yeast. A hybrid polypeptide consisting of hSOD and HIV1 *pol* amino acid sequences that encode the protease is accumulated within the cell when the aspartic acid located at the putative active site of the protease[56] was replaced by a glutamic acid residue (PR D25E). This observation provides new evidence that this aspartic acid residue is required for the catalytic activity of the HIV1 protease, as reported by others.[13–16]

In order to express the protease in a more purified state, a highly efficient secretion pathway in yeast was employed. By fusing the leader sequence of the yeast α-factor pheromone to either a 179 amino acid precursor of the protease or to the 99 amino acid mature form of the protease the viral enzyme was exported from the cell. This was verified by immunological detection of the viral protease in the media and by showing enzymatic activity of the secreted protease in vitro on the substrate $Pr53^{gag}$. Inefficient but detectable self-processing was observed when the active site variant PR179 D25E was secreted resulting in the accumulation of partially processed precursors, PRC and PRN. This suggests that the variant protease has reduced but measurable autocatalytic activity. However, the possibility that such polypeptides are generated by the enzymatic activity of a yeast protease cannot be excluded at this time. It is more likely that the glutamic acid can still act as a general acid/general base in the hydrolysis reaction but is inefficient when compared to an aspartic acid. The extra methylene group of the Glu presumably places the active site carboxylate out of catalytic register, resulting in a reduced activity of the protease. We are currently investigating whether secreted PR D25E can process the $Pr53^{gag}$ precursor and various synthetic substrates in vitro.

Secretion of the viral enzyme together with its uncommonly high isoelectric point and extremely hydrophobic character allowed rapid purification of the enzyme using preparative isoelectric focusing and reverse-phase HPLC. The media from transformed yeast was concentrated by ultrafiltration and the secreted protease was retained by a 10,000 molecular weight cut-off membrane. This is consistent with the predicted homodimeric quaternary structure of the protease of molecular weight 20,000. The protease has an isoelectric point (pI) of approximately 10 as judged by the preparative isoelectric focusing profile. This is in agreement with the predicted pI of the enzyme which was calculated from the individual pIs of the amino acids. Since the yeast secretion vector takes advantage of the natural proteolytic processing events that occur along the α-factor secretion pathway, this expression system can be used to overproduce variant forms of the protease that do not exhibit catalytic activity.

Gas phase sequence analysis of the purified protease resulted in a primary product that began at Pro1 of the mature enzyme. This result confirmed the predicted amino acid sequence that results from proteolysis by the yeast KEX2 processing enzyme.[57]

A sequence lacking the first five amino acids was also found by sequence analysis of the purified protease. This sequence can be generated by proteolysis of the Leu 5–Trp 6 peptide bond and may represent an autocatalytic process of the protease. The role this proteolysis product plays in the activity of the enzyme remains to be determined. The efficient yeast expression systems described for the intracellular or extracellular expression of the HIV1 protease will facilitate future biochemical studies on authentic or engineered variants of this enzyme. The vectors should also prove to be useful for other members of the aspartyl protease family associated with retroviral polyprotein processing. Finally, it may be possible to use yeast secretion of active viral protease to develop rapid in vitro screens and/or in vivo selections to monitor the efficiency of potential inhibitors of the aspartyl proteolytic activity.

ACKNOWLEDGMENTS

We gratefully acknowledge the contribution of Dr. Frank Masiarz for gas phase sequence analysis, Dr. Ian Bathurst for providing us with purified $Pr53^{gag}$, and Helen Gibson for assistance in the initial DNA constructions. This work was supported by Grant GM 39552 from the National Institute of Health. LMB was a recipient of a University of California Task Force Fellowship F88SF122.

REFERENCES

1. Wellink, J., van Kammen, A. Proteases involved in the processing of viral polyproteins. Arch. Virol. 98:1–26, 1988.
2. Wong-Staal, F., Gallo R.C. Human T-lymphotropic retroviruses. Nature (London) 317:395–403, 1985.
3. Coffin, J.M. In: "Molecular Biology of Tumor Viruses, RNA Tumor Viruses," Weiss, R., Teich, N., Varmus, H., Coffin, J., eds. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab., 261–368, 1982.
4. Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., Weiss, R. Human immunodeficiency viruses. Science 232:697, 1986.
5. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S.R., Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T. Gallo, R.C., Wong-Staal, F. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) 313:277–284, 1985.
6. Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D., Luciw, P.A. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science 227:484–492, 1985.
7. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., Alizon, M. Nucleotide sequence of the AIDS virus, LAV. Cell 40:9–17, 1985.
8. Kräusslich, H-G., Wimmer, E. Viral proteinases. Annu. Rev. Biochem. 57:701–754, 1988.
9. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M., Sigal, I.S. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. U.S.A. 85:4686–4690, 1988.
10. Debouck, C., Gorniak, J.G., Strickler, J.E., Meek, T.D., Metcalf, B.W., Rosenberg, M. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. Proc. Natl. Acad. Sci. U.S.A. 84:8903–8906, 1987.
11. Giam, C-Z., Boros, I. *In vivo* and *in vitro* autoprocessing of human immunodeficiency virus protease expressed in *Escherichia coli*. J. Biol. Chem. 263:14617–14620, 1988.
12. Hansen, J., Billich, S., Schulze, T., Sukrow, S., Moelling, K. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J. 7:1785–1791, 1988.
13. Le Grice, S.F.J., Mills, J., Mous, J. Active site mutagenesis of the AIDS virus protease and its alleviation by *trans* complementation. EMBO J. 7:2547–2553, 1988.
14. Loeb, D.D., Hutchison, C.A., III, Edgell, M.H., Farmerie, W.G., Swanstrom, R. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. J. Virol. 63:111–121, 1989.
15. Mous, J., Heimer, E.P., Le Grice, S.F.J. Processing protease and reverse transcriptase from human immunodeficiency virus type 1 polyprotein in *Escherichia coli*. J. Virol. 62:1433–1436, 1988.
16. Seelmeier, S., Schmidt, H., Turk, V., van der Helm, K. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. Proc. Natl. Acad. Sci. U.S.A. 85:6612–6616, 1988.
17. Navia, M.A., Fitzgerald, M.D., McKeever, B.M., Leu, C-T., Heimbach J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV1. Nature (London) 337:615–620, 1989.
18. Miller, M., Jaskólski, M., Rao, J.K.M., Leis, J., Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature (London) 337: 576–579, 1989.
19. Kramer, R.A., Schaber, M.D., Skalka, A.M., Ganguly, K., Wong-Staal, F., Reddy, E.P. HTLV-III *gag* protein is processed in yeast cells by the virus *pol*-protease. Science 231: 1580–1584, 1986.
20. Barr, P.J., Power, M.D., Lee-Ng, C.T., Gibson, H.L., Luciw, P. Expression of active human immunodeficiency virus reverse transcriptase in *Saccharomyces cerevisiae*. Bio/Technology 5:486–489, 1987.
21. Bitter, G.A., Chen, K.K., Banks, A.R., Lai, P.-H. Secretion of foreign proteins from *Saccharomyces cerevisiae* directed by alpha-factor gene fusions. Proc. Natl. Acad. Sci. U.S.A. 81:5330–5334, 1984.
22. Cousens, L.S., Shuster, J.R., Gallegos, C., Ku, L., Stempien, M.M., Urdea, M.S., Sanchez-Pescador, R., Taylor, A., Tekamp-Olson, P. High level expression of proinsulin in the yeast, *Saccharomyces cerevisiae*. Gene 61:265–275, 1987.
23. Hinnen, A., Hicks, J.B., Fink, G.R. Transformation of yeast. Proc. Natl. Acad. Sci. U.S.A. 75:1929–1933, 1978.
24. Travis, J., Owen, M., George, P., Carrell, R., Rosenberg, S., Hallewell, R.A., Barr, P.J. Isolation and properties of recombinant DNA produced variants of human alpha 1-proteinase inhibitor. J. Biol. Chem. 260:4384–4389, 1985.
25. Brake, A.J., Merryweather, J.P., Coit, D.G., Heberlein, U.A., Masiarz, F.R., Mullenbach, G.T., Urdea, M.S., Valenzuela, P., Barr, P.J. Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. 81:4642–4646, 1984.
26. Gardell, S.J., Craik, C.S., Hilvert, D., Urdea, M.S., Rutter, W.J. Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis. Nature (London) 317:551–555, 1985.
27. Singh, A., Lugovoy, J.M., Kohr, W.J., Perry, L.J. Synthesis, secretion and processing of α-factor-interferon fusion proteins in yeast. Nucl. Acids Res. 12:8927–8938, 1984.
28. Barr, P.J., Steimer, K.S., Sabin, E.A., Parkes, D., George-Nascimento, C., Stephans, J.C., Powers, M.A., Gyenes, A., Van Nest, G.A., Miller, E.T., Higgins, K.W., Luciw, P.A. Antigenicity and immunogenicity of domains of the human immunodeficiency virus (HIV) envelope polypeptide expressed in the yeast *Saccharomyces cerevisiae*. Vaccine 5:90–101, 1987.
29. Cregg, J.M., Tschopp, J.F., Stillman, C., Siegel, R., Akong, M., Craig, W.S., Buckholz, R.G., Madden, K.R., Kellaris, P.A., Davis, G.R., Smiley, B.L., Cruze, J., Torregrossa, R., Velicelebi, G., Thill, G.P. High-level expression and effi-

cient assembly of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. Bio/Technology 5: 479–485, 1987.
30. Sharma, S., Godson, G.N. Expression of the major surface antigen of *Plasmodium knowlesi* sporozoites in yeast. Science 228:879–882, 1985.
31. Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G., Hall, B.D. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. Nature (London) 298: 347–350, 1982.
32. Bathurst, I.C., Chester, N., Gibson, H.L., Dennis, A.F., Steimer, K.S., Barr, P.J. N-Myristylation of the human immunodeficiency virus type-I gag polyprotein precursor in *Saccharomyces cerevisiae*. J. Virol., 63:3176–3179, 1989.
33. Barr, P.J., Cousens, L.S., Lee-Ng, C.T., Medina-Selby, A., Masiarz, F.R. Hallewell, R.A., Chamberlain, S.H., Bradley, J.D., Lee, D., Steimer, K.S., Poulter, L., Burlingame, A.L., Esch, F., Baird, A. Expression and processing of biologically active fibroblast growth factors in the yeast *Saccharomyces cerevisiae*. J. Biol. Chem. 263:16471–16478, 1988.
34. Hallewell, R.A., Mills, R., Tekamp-Olson, P., Blacher, R., Rosenberg, S., Otting, F., Masiarz, F.R., Scandella, C.J. Amino terminal acetylation of authentic human Cu,Zn superoxide dismutase produced in yeast. Bio/Technology 5: 363–366, 1987.
35. Fauci, A.S. The human immunodeficiency virus: Infectivity and mechanism of pathogenesis. Science 239:617–622, 1988.
36. Piot, P., Plummer, F.A., Mhalu, F.S., Lamboray, J-L., Chin, J., Mann, J.M. AIDS: An international perspective. Science 239:573–579, 1988.
37. Luciw, P.A., Potter, S.J., Steimer, K., Dina, D., Levy, J.A. Molecular cloning of AIDS-associated retrovirus. Nature (London) 312:760–763, 1984.
38. Erhart, E., Hollenberg, C.P. The presence of a defective LEU2 gene on 2u DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. J. Bacteriol. 156:625–635, 1983.
39. Steimer, K.S., Higgins, K.W., Powers, M.A., Stephans, J.C., Gyenes, A., George-Nascimento, C., Luciw, P.A., Barr, P.J., Hallewell, R.A., Sanchez-Pescador, R. Recombinant polypeptide from the endonuclease region of the acquired immune deficiency syndrome retrovirus polymerase (pol) gene detects serum antibodies in most infected individuals. J. Virol. 58:9–16, 1986.
40. Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J., Struhl, K. In: "Current Protocols in Molecular Biology." New York: Wiley, 1989:131.
41. Evnin, L., Craik, C.S. Development of an efficient method for generating and screening active trypsins and variants. Ann. N.Y. Acad. Sci. 542:61–74, 1988.
42. Knowles, J.R. Tinkering with enzymes: What are we learning? Science 236:1252–1258, 1987.
43. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685, 1970.
44. The nomenclature adapted for the virally encoded proteins is that suggested by Leis, J., Baltimore, D., Bishop, J.M., Coffin, J., Fleissner, E., Goff, S.P., Oroszlan, S., Robinson, H., Skalka, A.M., Temin, H.M., Vogt, V. Standardized and simplified nomenclature for proteins common to all retroviruses. J. Virol 62:1808–1809, 1988. The nomenclature is as follows: *gag* polyprotein precursor, pr53^{gag}; matrix protein (MA, p17); capsid protein (CA, p25/24); nucleocapsid protein (NC, p9); protease (PR, p10); reverse transcriptase (RT, p66/51); integration protein (IN, p34); envelope surface protein (SU, gp120); and envelope transmembrane protein (TM, gp40).
45. Steimer, K.S., Puma, J.P., Power, M.D., Powers, M.A., George-Nascimento, C. Stephans, J.C. Levy, J.A., Sanchez-Pescador, R., Luciw, P.A., Barr, P.J., Hallewell, R.A. Differential antibody responses of individuals infected with AIDS-associated retroviruses surveyed using the viral core antigen p25gag expressed in bacteria. Virology 150:283–290, 1986.
46. Hunkapiller, M.W. "Applied Biosystem-User Bulletin Number 14." Foster City, California: Applied Biosystem, 1985.
47. LeGendre, N., Matsudaira, P. Direct protein microsequencing from immobilon-P transfer membrane. BioTechniques 6:154–159, 1988.
48. Babé, L.M., Pichuantes, S., Barr, P.J., Masiarz, F., Bathurst, I., Craik, C.S. HIV1 protease: Bacterial expression, purification and characterization. J. Cell. Biochem. 1989 (in press).
49. Graves, M.C., Lim, J.J., Heimer, E.P., Kramer, R.A. An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity. Proc. Natl. Acad. Sci. U.S.A. 85:2449–2453, 1988.
50. Leuthardt, A., Le Grice, S.F.J. Biosynthesis and analysis of a genetically engineered HIV-1 reverse transcriptase/endonuclease polyprotein in *Escherichia coli*. Gene 68:35–42, 1988.
51. Schultz, A.M., Oroszlan, S. *In vivo* modification of retroviral gag gene-encoded polyproteins by myristic acid. J. Virol. 46:355–361, 1983.
52. Henderson, L.E., Krutzsch, H.C., Oroszlan, S. Myristyl amino-terminal acylation of murine retrovirus proteins: An unusual post-translational protein modification. Proc. Natl. Acad. Sci. U.S.A. 80:339–343, 1983.
53. Rein, A., McClure, M.R., Rice, N.R., Luftig, R.B., Schultz, A.M. Myristylation site in Pr65^{gag} is essential for virus particle formation by Moloney murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. 83:7246–7250, 1986.
54. Mervis, R.J., Ahmad, N., Lillehoj, E.P., Raum, M.G., Salazar, F.H.R., Chan, H.W., Venkatesan, S. The gag gene products of human immunodeficiency virus type 1: Alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. J. Virol. 62:3993–4002, 1988.
55. Schneider, J., Kent, S.B.H. Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. Cell 54:363–368, 1988.
56. Pearl, L.H., Taylor, W.R. A structural model for the retroviral proteases. Nature (London) 329:351–354, 1987.
57. Julius, D., Brake, A., Blair, L., Kunisawa, R., Thorner, J. Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-α-factor. Cell 37:1075–1089, 1984.
58. di Marzo Veronese, F., Copeland, T.D., Oroszlan, S., Gallo, R.C., Sarngadharan, M.G. Biochemical and immunological analysis of human immunodeficiency virus *gag* gene products p17 and p24. J. Virol. 62:795–801, 1988.
59. Lightfoote, M.M., Coligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A., Venkatesan, S. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retroviruses. J. Virol. 60:771–775, 1986.
60. Lillehoj, E.P., Salazar, F.H.R., Mervis, R.J., Raum, M.G., Chan, H.W., Ahmad, N., Venkatesan, S. Purification and structural characterization of the putative *gag-pol* protease of human immunodeficiency virus. J. Virol. 62: 3053–3058, 1988.
61. Nutt, R.F., Brady, S.F., Darke, P.L., Ciccarone, T.M., Colton, C.D., Nutt, E.M., Rodkey, J.A., Bennett, C.D., Waxman, L.H., Sigal, I.S., Anderson, P.S., Veber, D.F. Chemical synthesis and enzymatic activity of a 99-residue peptide with a sequence proposed for the human immunodeficiency virus protease. Proc. Natl. Acad. Sci. U.S.A. 85: 7129–7133, 1988.

PROTEINS

Structures and Molecular Principles

THOMAS E. CREIGHTON
Medical Research Council
Laboratory of Molecular Biology
Cambridge, England

W. H. FREEMAN AND COMPANY NEW YORK

SDS Polyacrylamide Gel Electrophoresis

A related technique is **polyacrylamide gel electrophoresis,** wherein the rate of migration through a gel with small pores is dependent upon the size of the protein. As with the other techniques discussed, the shape of the protein will also determine the rate of migration. An additional factor in this case is the net charge of the protein, since electrophoresis is the driving force. These two factors may be determined by varying the sieving effect of the gel, achieved by altering the concentration of acrylamide and the cross-linking agent, methylene-*bis*-acrylamide; all other parameters are kept constant so that the shape and net charge do not vary.

A much more satisfactory approach is to abolish all shape and charge differences of proteins by adding the denaturant sodium dodecyl sulfate (SDS):

$$CH_3—(CH_2)_{11}—SO_4^-Na^+ \qquad (1–58)$$

This approach has arisen from a series of empirical observations, so that its detailed physical basis is not yet fully understood. Binding studies of a variety of different proteins have shown that above an SDS monomer concentration of 8×10^{-4} M, a constant of 1.4 grams of SDS is bound per gram of protein, i.e., one molecule of SDS for every two amino acid residues of the chain. This high level of binding of the charged detergent and the constant binding ratio will generally "swamp out" the intrinsic charge contribution of most proteins, so that an approximately constant negative charge per unit mass will be obtained. All polypeptides also appear to have a similar shape when SDS is bound, generally considered as elongated particles, with a constant diameter and a length proportional to the number of amino acid residues in the polypeptide chain. The exact nature of the protein–SDS complex is not known; none of the models proposed is entirely consistent with the many experimental observations.

Nevertheless, the constant net charge and shape lead to SDS–protein complexes having electrophoretic mobilities in polyacrylamide gels that are generally directly proportional to the logarithm of the length of the polypeptide chain (Figure I–5). By comparing the mobility of an unknown protein with that of a set of standard marker proteins, the molecular weight of the unknown polypeptide chain may usually be determined within 10 per cent of the true value. In certain cases, abnormalities in SDS binding or protein conformation, large differences in intrinsic protein charge, or covalently attached nonprotein moieties may lead to increased or decreased electrophoretic mobilities; therefore, caution is advisable in use of this technique. Nevertheless, the remarkable resolution and ease of SDS electrophoresis has made it the most widely used method for determining the absolute values of polypeptide chain sizes. It is especially useful in studies of processes in which the molecular weights of proteins are altered.

Molecular weight estimation of polypeptide chains by electrophoresis in SDS–polyacrylamide gels. A. L. Shapiro, et al. Biochem. Biophys. Res. Commun. 28:815–820, 1967.

Introduction to Protein Structure

Carl Branden
Uppsala Biomedical Center
Swedish University of Agricultural Sciences
Uppsala, Sweden

John Tooze
European Molecular Biology Organization
Heidelberg, Germany

Garland Publishing, Inc.
New York and London, 1991

THE COVER

Front: The background photograph of the cover is of a Laue x-ray diffraction pattern produced by a crystal of the plant enzyme ribulose bisphosphate carboxylase. This technique is described in Chapter 17. Information derived from such x-ray patterns, together with a knowledge of the amino acid sequence, enabled the three-dimensional arrangement of atoms in the protein to be determined. A simplified representation of this protein structure is shown in color, superimposed on the diffraction pattern. The enzyme, which is involved in the fixation of carbon dioxide, is a member of the large class of α/β barrel protein structures. This class of structures is discussed in detail in Chapter 4.

Back: Tomato bushy stunt virus is a spherical virus made from 180 protein subunits. Arms extending from sixty of these subunits contribute to an internal framework that determines the size of the correctly assembled virus particle. The interdigitated arms from three subunits meet at each of the twenty icosahedral threefold axes of the virus. One such axis is shown here with the β strands from three subunits shown in different shades of green. Virus structure is described in more detail in Chapter 11.

Library of Congress Cataloging-in-Publication Data

Branden, Carl.
Introduction to protein structure / Carl Branden, John Tooze.
p. cm.
Includes index.
ISBN 0-8153-0344-0 — ISBN 0-8153-0270-3 (pbk.)
1. Proteins—Structure. I. Tooze, John. II. Title.
QP551.B7635 1991
574.19'245—dc20 91-11788
CIP

Published by Garland Publishing, Inc.
136 Madison Ave., New York, New York, 10016

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

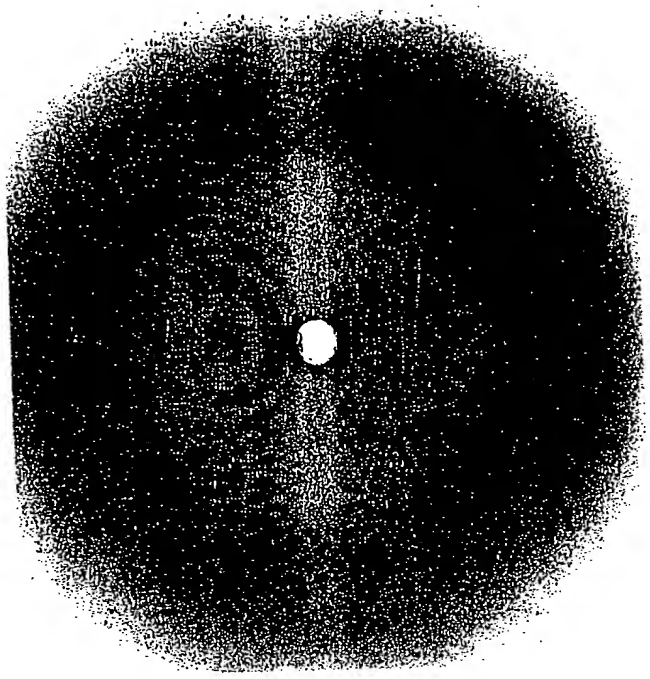


Figure 13.4 X-ray diffraction pattern from crystals of a membrane-bound protein, the bacterial photosynthetic reaction center. (Courtesy of Hartmut Michel.)

many, displayed the x-ray diagram shown in Figure 13.4. Not only was this the first x-ray picture to high resolution of a membrane protein, but the crystal was formed not from a small protein of trivial function but from a large complex of polypeptide chains that represents a class of proteins having a function of central importance for life on earth. The protein complex was a **photosynthetic reaction center** from the photosynthetic bacterium *Rhodopseudomonas viridis*, which converts the energy of captured sunlight into electrical and chemical energy in the first steps of photosynthesis. The structure has subsequently been solved to 2.5 Å resolution by H. Michel in collaboration with Hans Deisenhofer and Robert Huber at the same institute.

The interiors of *Rhodopseudomonad* bacteria are filled with photosynthetic vesicles, which are hollow membrane-enveloped spheres. The photosynthetic reaction centers are embedded in the membrane of these vesicles. One end of the protein complex faces the inside of the vesicles, which is known as the periplasmic side, the other end faces the cytoplasm of the cell. Around each reaction center, on the periplasmic side within the vesicles, there are hundreds of small membrane proteins, the antenna pigment protein molecules with bound chlorophyll. These catch photons over a wide area and funnel them to the reaction center. By this arrangement the reaction center can utilize about a hundred times more photons than those that directly strike the special pair of chlorophyll molecules at the heart of the reaction center.

The reaction center is built up from four polypeptide chains, three of which are called L, M, and H because they have light, medium, and heavy molecular masses as deduced from their electrophoretic mobility on SDS-PAGE. Subsequent amino acid sequence determinations showed, however, that the H chain is in fact the smallest with 258 amino acids, followed by the L chain with 273 amino acids. The M chain is the largest polypeptide with 323 amino acids. This discrepancy between apparent relative masses and real molecular weights underlines the uncertainty in deducing molecular masses of membrane-bound proteins from their mobility in electrophoretic gels.

The L and M subunits show sequence identity of about 25% and are therefore homologous and evolutionarily related proteins. The H subunit, on the other hand, has a completely different sequence. The fourth subunit of the reaction center is a cytochrome that has 336 amino acids with a sequence that is not homologous to any other known cytochrome sequence.

In addition to these polypeptide chains, the reaction center contains a number of pigments. There are four bacteriochlorophyll molecules (Figure

JOURNAL OF VIROLOGY, Dec. 2004, p. 13708–13716 Vol. 78, No. 24
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.24.13708–13716.2004

Identification of Residues in the Dengue Virus Type 2 NS2B Cofactor That Are Critical for NS3 Protease Activation

Pornwaratt Niyomrattanakit, Pakorn Winoyanuwattikun, Santad Chanprapaph, Chanan Angsuthanasombat, Sakol Panyim, and Gerd Katzenmeier*

Laboratory of Molecular Virology, Institute of Molecular Biology and Genetics, Mahidol University, Nakompathom, Thailand

Received 26 February 2004/Accepted 8 August 2004

Proteolytic processing of the dengue virus polyprotein is mediated by host cell proteases and the virus-encoded NS2B-NS3 two-component protease. The NS3 protease represents an attractive target for the development of antiviral inhibitors. The three-dimensional structure of the NS3 protease domain has been determined, but the structural determinants necessary for activation of the enzyme by the NS2B cofactor have been characterized only to a limited extent. To test a possible functional role of the recently proposed $\Phi x_3 \Phi$ motif in NS3 protease activation, we targeted six residues within the NS2B cofactor by site-specific mutagenesis. Residues Trp62, Ser71, Leu75, Ile77, Thr78, and Ile79 in NS2B were replaced with alanine, and in addition, an L75A/I79A double mutant was generated. The effects of these mutations on the activity of the NS2B(H)-NS3pro protease were analyzed in vitro by sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of autoproteolytic cleavage at the NS2B/NS3 site and by assay of the enzyme with the fluorogenic peptide substrate GRR-AMC. Compared to the wild type, the L75A, I77A, and I79A mutants demonstrated inefficient autoproteolysis, whereas in the W62A and the L75A/I79A mutants self-cleavage appeared to be almost completely abolished. With exception of the S71A mutant, which had a k_{cat}/K_m value for the GRR-AMC peptide similar to that of the wild type, all other mutants exhibited drastically reduced k_{cat} values. These results indicate a pivotal function of conserved residues Trp62, Leu75, and Ile79 in the NS2B cofactor in the structural activation of the dengue virus NS3 serine protease.

Infection by dengue viruses is now widely recognized as a major public health concern, with more than 1 million cases of dengue hemorrhagic fever per year and case fatality rates ranging from 1 to 10% (23). There are four serotypes of dengue virus, which cause dengue hemorrhagic fever and dengue shock syndrome (21, 22). Dengue viruses, members of the *Flaviviridae* family, are small, enveloped, positive-stranded RNA viruses which are transmitted by *Aedes* mosquitoes (7). At present, neither a commercial vaccine nor a causative treatment is available for the prevention or cure of acute dengue virus diseases.

The genomic RNA of dengue virus serotype 2 contains 10,723 nucleotides and encodes a large polyprotein precursor of 3,391 amino acid residues which consists of three structural proteins (C, prM, and E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (26). The polyprotein is co- and posttranslationally processed by proteases of the host cell and the virus-encoded two-component protease NS2B-NS3 to generate individual viral proteins (11, 18). Optimal activity of the NS3 serine protease (flavivirin, EC 3.4.21.91) is an essential requirement for maturation of the virus, and inhibition of this enzyme offers the prospect of an effective antiviral chemotherapy for severe cases of dengue hemorrhagic fever and dengue shock syndrome (for review see references 6, 39, and 41 and references herein).

* Corresponding author. Mailing address: Laboratory of Molecular Virology, Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus, Phutthamonthon 4 Rd., Nakornpathom 73170, Thailand. Phone: (662)-800-3624-1237. Fax: (662)-441-9906. E-mail: frkgz@mahidol.ac.th.

The NS2B-NS3 two-component protease mediates cleavage in the nonstructural region of the viral polyprotein at the NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5 junctions. Additional cleavages within the C, NS2A, and NS4A proteins and within a C-terminal portion of NS3 itself were described in the literature (1, 34, 35, 46). With the exception of the NS2B/NS3 junction, which contains a glutamine residue at the P2 position, the cleavage sites for the NS3 protease consist of a pair of dibasic amino acids (Arg or Lys) at the P1 and P2 positions, followed by a short-chain amino acid (Gly, Ala, or Ser) at the P1′ site (12, 13, 40, 48, 51). The minimum domain size required for protease activity of the 69-kDa NS3 protein has been mapped to 167 residues at the N terminus (33). Based on sequence comparisons with known serine proteases, a catalytic triad comprised of residues His51, Asp75, and Ser135 was identified, and replacement of the catalytic Ser135 residue by alanine resulted in an enzymatically inactive NS3 protease (47). The C-terminal two-thirds of the dengue virus NS3 protein are associated with the enzymatic functions of a nucleoside triphosphatase and RNA helicase (20, 27). The three-dimensional structure of the NS3 protease domain (NS3pro) encompassing the N-terminal 185 amino acids has been resolved by X-ray crystallography, and the protein exhibits the six-stranded β-barrel conformation typical of chymotrypsin-like serine proteases (31).

The presence of a small activating protein or cofactor is a prerequisite for optimal catalytic activity of the flaviviral proteases with natural polyprotein substrates (4, 13). Although the dengue virus NS3 protease exhibits NS2B-independent activity with model substrates for serine proteases such as *N*-α-benzo-

yl-L-arginine-*p*-nitroanilide, enzymatic cleavage of dibasic peptides is markedly enhanced with the NS2B-NS3 cocomplex, and the presence of the NS2B cofactor was shown to be an absolute requirement for *trans* cleavage of a cloned polyprotein substrate (50). Intramolecular cleavage at the NS2B/NS3 site conducive to the formation of a noncovalent complex was observed with the NS2B(H)-NS3pro molecule after purification from overexpressing *Escherichia coli* and subsequent refolding (50).

A genetically engineered NS2B(H)-NS3pro protease containing a noncleavable nonamer glycine linker between the NS2B activation sequence and the protease moiety exhibited higher specific activity with *para*-nitroanilide peptide substrates than the NS2B(H)-NS3pro molecule (32). Recently we have shown that the NS2B-NS3pro protease incorporating a full-length NS2B cofactor sequence could catalyze the cleavage of 12-mer peptide substrates representing native polyprotein junctions (28, 29). However, this protein appeared to be completely resistant to proteolytic self-cleavage.

The initial characterization of the cofactor requirement for the dengue virus NS3 protease revealed that the minimal region required for protease activity was located in a 40-residue central hydrophilic segment of NS2B spanning residues Leu54 to Glu93 (19). Mutagenesis experiments with the yellow fever virus NS2B protein demonstrated that specific residues within this core sequence are critical for protease activation (14). Deletion of residues 51 to 55, 53 to 55, and 56 to 93 within the conserved central domain yielded no detectable processing of an NS2B-NS3pro polyprotein precursor, whereas a four-amino-acid deletion of the sequence ^{67}ISGS^{70} generated a protease with significantly reduced cleavage efficiency. Directed mutagenesis within the yellow fever virus NS2B protein confirmed a structural role for the N-terminal region of the conserved cofactor segment (17). Mutations within a charged N-terminal cluster comprising residues ^{52}ELKK^{55} impaired *cis* cleavage activity at the NS2B/NS3 site, and deletion analysis revealed that the conserved domain alone provided only basal cofactor activity, while the optimal function of the cofactor required both hydrophobic flanking regions of NS2B.

A significant reduction of NS3 cleavage activity was observed for the alanine substitutions at residues Val95 and Gln96 within the dengue virus NS3 protease sequence (37). It was proposed that these two residues are located at the C terminus of the NS2B binding cleft and that they are involved in precleavage association of NS2B with NS3 and proper processing at the NS2B/NS3 site.

An essential requirement for the activation of the protease is the presence of hydrophobic residues in the cofactor, which may act as an anchor in the enzyme-cofactor complex. Within the hepatitis C virus NS4A cofactor, two residues, Ile25 and Ile29, are critical for complete activation of the NS3 protease (9, 44, 45). For the NS4A cofactor from GB virus, a minimum region which supports NS3 protease activity was mapped to a sequence spanning residues Phe22 to Val36, and two central residues, Val27 and Trp31, were indispensable for maximal proteolytic activity (10).

A peptide comprising residues Ser69 to Glu81 of the dengue virus NS2B cofactor was recently proposed as a substitute for the cofactor (8), however, it failed to reconstitute catalytic activity of the NS3pro protease in vitro (32). Therefore, it seems likely that additional residues located further at the N terminus of the NS2B core sequence play a role in NS3 activation.

These findings were suggestive of a common structural motif involved in activation of the flaviviral proteases. The $\Phi x_3 \Phi$ motif is comprised of two bulky hydrophobic residues separated by three unspecified residues, and it was speculated that additional residues located outside this sequence motif would contribute to the stringent specificity of the protease for the corresponding polyprotein substrate (10).

Substantial biochemical data were accumulated for the hepatitis C virus protease which offer some structural and mechanistic explanations for the activation of this flaviviral protease by its cofactor. Interaction of the NS3 protease with the NS4A cofactor was shown to affect the folding of the NS3 protease, resulting in conformational rearrangements of the N-terminal 28 residues of the protease and a strand displacement conducive to the formation of a well-ordered array of three β-sheets in which the cofactor becomes an integral part of the protease fold (30, 49). The result of these conformational changes is a reorientation of the residues of the catalytic triad which is more favorable for proton shuttling during catalysis. A number of studies provided evidence that structural rearrangements leading to a fully activated protease are induced not only by binding of the cofactor but also by the substrate, as shown for the competitive inhibition of the NS3 protease by its cleavage products (2, 3, 24, 25, 38). Based on these findings, the hepatitis C virus enzyme has been described as an induced-fit protease (6, 39). However, analogies to the hepatitis C virus system should be treated with caution, since preliminary data obtained with the dengue virus and the GB virus enzymes are indicative of major structural differences in the activation process (10, 50).

We demonstrate in this report that alanine substitutions at residues Trp62, Leu75, and Ile79 in the dengue virus NS2B cofactor result in marked effects on autoprocessing at the NS2B/NS3 site and that activity of the mutant NS3 proteases with the synthetic peptide substrate is mainly affected by significantly reduced k_{cat} values. To analyze the structure-activity relationships which we have observed experimentally, we generated a molecular model for the NS2B/NS3 cocomplex based on homology to the hepatitis C virus NS3/NS4A protease.

MATERIALS AND METHODS

Construction of pTH/NS2B(H)-NS3pro by SOE-PCR. The recombinant plasmid encoding the dengue virus serotype 2 NS2B(H)-NS3pro protein was generated with the previously described plasmid pTH/NS2B-NS3 as a template for splicing by overlap extension (SOE)-PCR (15). The sequence for NS2B(H) was obtained with primers 5′-TGCTCACTGGAGGATCCGCCGATTTGGAACT GGAG-3′ (nucleotides 4259 to 4293 in dengue virus serotype 2) and 5′-CTTC ACTTCCCACAGGTACCACAGTGTTTGTTCTTCCTC-3′ (nucleotides 4399 to 4416). For amplification of NS3pro, primers 5′-GAACAAACACTGTGGTA CCTGTGGGAAGTGAAGAAAC-3′ (nucleotides 4492 to 4516) and 5′-CTTC TCTTTCAGGATCCCTAATCTTCGATCTCTGGGTTG-3′ (nucleotides 5043 to 5081) were used.

SOE-PCR was performed with a combination of both templates with the NS2B(H) forward and the NS3pro reverse primers incorporating an overlapping region of 33 nucleotides. The product of SOE-PCR comprises the sequence of NS2B from amino acid residues 48 to 95 followed by residues 121 to 130 and 180 N-terminal residues of the NS3 protease domain. The PCR product was cut with BamHI and cloned into the pTrcHisA expression vector (Invitrogen) to yield the polyhistidine-tagged fusion protein. The sequence of the resulting construct was

verified by DNA sequencing with an ABI Prism model 377 sequencer with a dye terminator cycle sequencing reaction kit (Perkin Elmer).

Mutation constructs. Alanine substitutions were introduced in the NS2B sequence at residues Trp62, Ser71, Leu75, Ile77, Thr78, and Ile79 with the QuikChange site-directed mutagenesis kit (Stratagene) following the manufacturer's directions. In addition, a double Leu75/Ile79 mutant was generated with the L75A mutant as the template for PCR with Ile79 mutagenic oligonucleotides. Additional base changes creating restriction sites suitable for screening of the resulting mutant constructs were introduced in the primer sequences.

The following pairs of forward and reverse primers were used for mutagenesis (bold letters indicate changed nucleotides, and italic letters represent restriction sites): W62A, 5′-CGATGTCA*AAGCT*GAAGACCAGGCAGAG-3′ and 5′-CTCTGCCTGGTCTTC*AGCTT*TGACATCG-3′; S71A, 5′-GAGATATCAGG*AG CTA*GTCCAATCC-3′ and 5′-GGATTGGAC*TAGCTC*CTGATATCTC-3′; L75A, 5′-GCAGTCC*GATCGC*GTCAATAACAATATCAG-3′ and 5′-CTGATATTGTTATTGACGC*GATC*GGACTGC-3′ ; I77A, 5′-CCAATCCTGTC*AGC TA*CAATCTCAGAAGATGG-3′ and 5′-CCATCTTCTGATATTGTAGC*TGA* CAGGATTGG-3′ ; T78A, 5′-CCAATCCTGTCAATAGCAATC*TCAGA*AGA TGG-3′ and 5′-CCATCTTCTGA*GA*TTGCTATTGACAGGATTGG-3′ ; and I79A, 5′-TCAATAACA*GCCTCAG*AAGATGGTAGC-3′ and 5′-GCTACCAT CTT*CTGAGGC*TGTTATTGAC-3′.

The catalytically inactive NS3 protease mutant with an S135A substitution was obtained as described earlier (28). Plasmid DNA from the mutants was analyzed by DNA sequencing to confirm that only the desired mutation was incorporated.

Expression and purification of protease constructs. The pTrcHis plasmids containing the recombinant NS2B(H)-NS3pro sequences were transformed into *Escherichia coli* C41(DE3). Transformants were grown in Luria broth (LB) medium supplemented with ampicillin (100 μg/ml) at 37°C. At an A_{600} of 0.5, isopropyl-1-thio-β-D-galactopyranoside was added to 0.1 mM, and the culture was grown at 37°C for 8 h. Cells were harvested by centrifugation (5,000 × *g*, 10 min, 4°C), resuspended in 20 ml of lysis buffer A (100 mM Tris-HCl, pH 7.5, 300 mM NaCl), and lysed with a French pressure cell at 14,000 lb/in^2. The lysate was clarified by centrifugation (10,000 × *g*, 30 min, 4°C), and the pellet fraction was washed two times with lysis buffer containing 1% Triton X-100.

Inclusion bodies were suspended in 15 ml of denaturing buffer B (100 mM Tris-HCl, pH 8.0, 300 mM NaCl, 8 M urea) followed by sonication (10 bursts at power setting 3 for 15 s) with a ultrasonic processor (Misonix). The suspension was centrifuged (10,000 × *g*, 30 min, 4°C), and the supernatant was loaded on a Hitrap chelating column (Pharmacia) equilibrated with denaturing buffer. The column was washed with 10 column volumes of denaturing buffer containing 20 mM imidazole and eluted at a flow rate of 0.5 ml min^{-1} with denaturing buffer containing 50 mM imidazole. Fractions of 1 ml were collected, and aliquots were analyzed for the presence of NS2B(H)-NS3pro by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) on 15% polyacrylamide gels.

Peak fractions were pooled and loaded on a Superdex 200 HR 10/30 gel filtration column (Pharmacia). The column was eluted with denaturing buffer at a flow rate of 0.3 ml min^{-1} and the fractions containing NS2B(H)-NS3pro, as analyzed by SDS-PAGE, were pooled and diluted with the same buffer to 0.5 mg ml^{-1}. Refolding of the protein was initiated by stepwise dialysis of 1-ml samples with a dialysis tubing (cutoff, 8 kDa) at 4°C against three changes of 100 mM Tris-HCl, pH 8.0–300 mM NaCl (200 ml) and one change against 200 ml of 100 mM Tris-HCl, pH 9.0–50 mM NaCl (buffer C). The dialysate was centrifuged (10,000 × *g*, 10 min, 4°C), and the protein concentration was determined with a Bradford protein assay kit (Bio-Rad). Preparations of the NS2B(H)-NS3pro protein were stored at −20°C in 100 mM Tris-HCl, pH 9.0–50 mM NaCl–50% glycerol.

Determination of NS3 protease activity. Autocleavage of NS2B(H)-NS3pro at the NS2B/NS3 site was monitored by Tricine-SDS-PAGE (43). Samples containing 0.5 μg μl^{-1} of purified NS2B(H)-NS3pro in buffer C were incubated at 37°C, and aliquots of 20 μl were removed at fixed intervals. The reaction was quenched by the addition of 7 μl of SDS-PAGE sample buffer (200 mM Tris-HCl, pH 7.5, 4% [wt/vol] SDS, 40% glycerol, 0.1% bromophenol blue, 100 mM dithiothreitol), and precursor processing was quantitated by densitometric analysis of band intensities obtained from SDS-PAGE with the ONE-D scan program (Scanalytics). Cleavage at the NS2B/NS3 site was confirmed by automated Edman amino acid sequencing of the protease fragment NS3pro and by Western blot analysis with anti-Xpress antibodies (Invitrogen) of the N-terminal cleavage fragment (His)$_6$NS2B(H).

The fluorogenic substrate GRR-AMC (Peptide International) was used for the in vitro assay of NS3. The assay was performed in 96-well microtiter plates with a Labsystems Fluoroscan II (Labsystems) at an excitation wavelength of 355 nm and emission wavelength of 460 nm at 37°C. Assay reactions contained in a 100-μl final volume NS2B(H)-NS3pro and the mutant proteins at a concentration of 0.15 μM in 100 mM Tris-HCl, pH 9.0. The substrate concentration was varied between 10 and 250 μM, and signals were converted to concentrations by comparison with standard amounts of free AMC. K_m and V_{max} values were obtained from measurements of initial velocities prior to 10% substrate depletion, assuming that 100% of the protease was enzymatically active. To determine K_m and V_{max} values, Michaelis-Menten kinetics, $v = V_{max}[S]/[S] + K_m$, were transformed into double reciprocal Lineweaver-Burk plots by nonlinear regression analysis with the GraphPad Prism software. Three independent experiments were carried out for each set of data points, and data are reported as mean ± standard error.

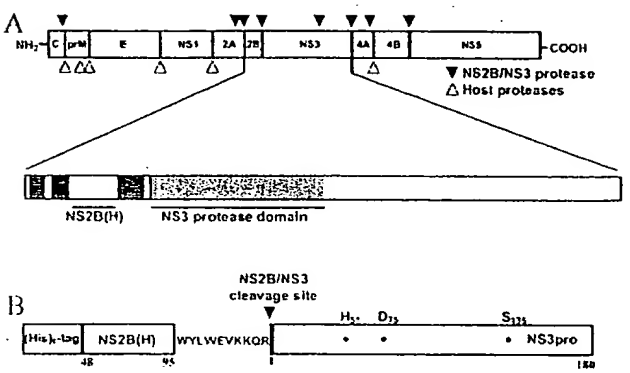


FIG. 1. Organization of the dengue virus polyprotein and the NS2B(H)-NS3(pro) construct. (A) Sites on the dengue virus polyprotein cleaved by host-encoded proteases (▽) and the virus-encoded two-component protease NS2B-NS3 (▼). The NS3 protease domain (NS3pro) is shaded, and the 40-residue minimum cofactor region which supports catalytic activity of the NS3 protease is indicated by a bar. (B) Structural organization of the expression construct NS2B(H)-NS3pro as generated by SOE-PCR. The construct contains the N-terminal polyhistidine tag for purification purposes, the central activation sequence of NS2B from residues 48 to 95 followed by residues 121 to 130, and 180 N-terminal residues of NS3 representing the protease domain. The residues of the catalytic triad, His51, Asp75, and Ser135, are indicated, and the sequence of the native NS2B/NS3 polyprotein cleavage junction is shown.

Coordinates and molecular modeling. The crystal structures of the dengue virus type 2 NS3 serine protease (protein database identifier 1BEF) (31) and the hepatitis C virus NS3/NS4A complex (protein database identifier 1JXP) (49) were obtained from the protein database (Brookhaven National Laboratory). To obtain a structure for the dengue virus NS2B core segment in the complex with the NS3 protease, the NS2B peptide (residues 56 to 93) was initially aligned with the NS4A peptide (residues 21 to 32) according to the published sequence comparison (10). The structure for the NS2B peptide was generated with the Modeller 6v2 software (42). The dengue virus serotype 2 protease domain was superimposed on the hepatitis C virus protease domain, and the structures of the NS3 protease and the NS2B peptide were combined. Several 1-ns molecular dynamics trajectories of NS3pro in complex with NS2B were generated. The simulation of the resulting structure was performed with the Gromacs simulation package (http://www.gromacs.org) with GROMOS96 force field in a single-point charge water box. The space between protein and box walls was set to a minimum distance of 7 Å, and the system was energy minimized with the steepest descent. The simulated structures were visualized by Deepview Swiss-PdbViewer v3.5b4. The final model was evaluated by a Ramachandran plot, and 98.2% of the nonglycine residues were in allowed conformations.

RESULTS

Generation of mutants. The sequence ^{75}LSITI^{79} represents the $\Phi x_3 \Phi$ motif in the NS2B cofactor of dengue virus serotype 2 which was recently proposed to play a functional role in the association of the flaviviral proteases with their corresponding cofactors (10). To analyze the effects of amino acid substitutions in the NS2B cofactor on the enzymatic activity of the NS3 serine protease, we constructed the NS2B(H)-NS3pro polypro-

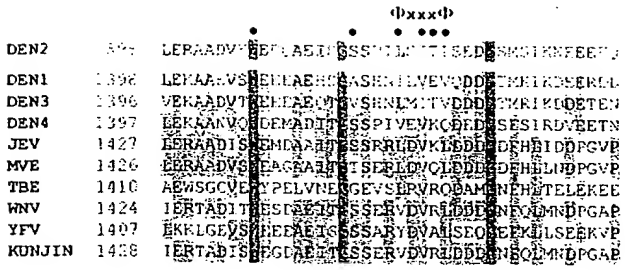


FIG. 2. Sequence alignment of the conserved central domain within the NS2B cofactors of members of the *Flaviviridae* family. Numbers on the left indicate positions of amino acid residues in the polyprotein sequence. Residues which are identical in all sequences are shaded in dark grey, and conserved residues are shaded in light grey. The location of the $\Phi x_3 \Phi$ motif is indicated above the alignment, and residues within the dengue virus NS2B cofactor that were changed to alanine are labeled by dots. Abbreviations: DEN, dengue virus; JEV, Japanese encephalitis virus; KUNJIN, Kunjin virus; MVE, Murray valley encephalitis virus; TBE, tick-borne encephalitis virus; WNV, West Nile virus; YFV, yellow fever virus.

tein precursor by SOE-PCR (Fig. 1) (15). A sequence alignment of known flavivirus cofactor sequences and the location of the $\Phi x_3 \Phi$ motif are shown in Fig. 2.

Previous work had shown that the NS2B(H)-NS3pro protein undergoes proteolytic self-cleavage at the NS2B/NS3 site that is conducive to the formation of a noncovalent complex (50). Alanine substitutions were introduced at residues Trp62, Ser71, Leu75, Ile77, Thr78, Ile79, and Leu75 plus Ile79 in the NS2B sequence. An enzymatically inactive NS2B(H)-NS3pro protein was obtained by replacing the catalytic Ser135 residue with alanine and used as negative control for the activity assays as described previously (28). All recombinant plasmids were subjected to DNA sequencing, and no mutations were found at nontargeted sites. Expression of the mutant derivatives as inclusion bodies in *E. coli* and purification under denaturing conditions by immobilized metal chelate chromatography and gel filtration yielded homogeneous products, as determined by SDS-PAGE analysis. The 29.8-kDa $(His)_6$-NS2B(H)-NS3pro molecule displayed anomalous migration in gel electrophoresis with a higher apparent molecular mass of approximately 37 kDa. Subsequent refolding was performed by stepwise dialysis, and correct cleavage at the NS2B/NS3 junction was confirmed for the wild-type protein by N-terminal amino acid sequencing of the 20-kDa cleavage product, which yielded the sequence AGVLW, identical to the first five residues of the NS3 protein.

Effect of alanine substitutions on self-cleavage efficiency. Protein samples purified by metal affinity chromatography and gel filtration were analyzed by SDS-PAGE for autocleavage at the NS2B/NS3 site after various periods of incubation (Fig. 3). After extensive dialysis, wild-type NS2B(H)-NS3pro exhibited complete autoproteolytic cleavage, which resulted in two protein products of 20 kDa and 10 kDa, whereas the S135A mutant was completely inactive in the self-cleavage assay. In Western blot analysis, only the 10-kDa protein reacted with anti-Xpress antibodies directed against the polyhistidine tag, which confirmed that this protein represents the N-terminal cleavage fragment $(His)_6$-NS2B(H) (data not shown).

The NS2B mutants displayed different levels of proteolytic processing when analyzed by SDS-PAGE immediately after refolding. A densitometric analysis based on the amount of the NS2B(H)-NS3pro precursor remaining revealed that self-cleavage was almost completely abolished in the L75A/I79A double mutant. Autoprocessing was markedly reduced in the W62A mutant, which gave approximately 10% of wild-type activity. In contrast, the cleavage efficiencies of the S71A and the T78A mutants were not significantly affected by the substitutions and were comparable to that of the wild type. A sequence alignment of known flaviviral cofactor sequences shows that a serine at position 71 is preferred in most viruses of the *Flaviviridae* family, and substitution of this residue with alanine did not have a marked effect on proteolytic self-cleavage.

Thr78 is part of the $\Phi x_3 \Phi$ motif, but this residue is not very well conserved among the *Flaviviridae* and we expected that one alanine residue in the context of the serotype 2 sequence would have only a marginal effect on the activity of the NS3 protease. In accordance with this prediction, at this position the presence of the hydrophobic alanine residue is well tolerated. In contrast, alanine replacements at the hydrophobic residues Leu75, Ile77, and Ile79 caused a reduction in autocleavage activity of the NS2B(H)-NS3pro protein. Substitutions at Leu75 and Ile79 reduced autoprocessing to approximately 55 and 75% of the wild-type value, and the L75A mutation had a greater effect on cleavage efficiency than the I77A substitution, which still allowed for approximately 80% of wild-type precursor cleavage.

The data presented here support an important function for the $\Phi x_3 \Phi$ motif in activation of the NS3 protease. In agreement with a critical role for this hydrophobic sequence element, we found that self-cleavage was substantially decreased for the L75A/I79A double mutant, which gave only 2% of wild-type cleavage.

We also examined the effect of an alanine replacement at the W62 residue. This position is strictly conserved among the members of the *Flaviviridae* family (Fig. 2) and is located in the N-terminal region of the NS2B activation sequence. Alanine substitution at this position had a dramatic effect on protease activity, and self-cleavage was markedly reduced with this mutant protein; a finding which suggests a pivotal function for this invariant residue in protease activation.

Delayed processing kinetics of the mutants. To answer the question of whether the alanine substitutions at critical positions within the NS2B cofactor resulted in a catalytically inefficient NS3 serine protease, we analyzed the levels of self-cleavage after various periods of incubation ranging from 1 to 24 h (Fig. 3).

The wild-type NS2B(H)-NS3pro molecule and the S71A mutant underwent complete autoproteolytic cleavage during the refolding process, whereas progressive proteolysis leading to complete cleavage of the precursor was observed for the I77A, T78A, and I79A mutants. Continued proteolysis of the S71A mutant protein resulted in additional cleavage fragments in SDS-PAGE at molecular masses of approximately 16 and 12 kDa. These proteins could represent fragments generated by cleavage at internal sequences within the NS3pro molecule. The NS3pro sequence contains paired basic residues at positions ^{63}KRI^{65} and ^{142}KKG^{144} and a monobasic site resembling the NS2B/NS3 junction at ^{27}QRG^{29} which could serve as additional substrates for the protease, and cleavage at these sites

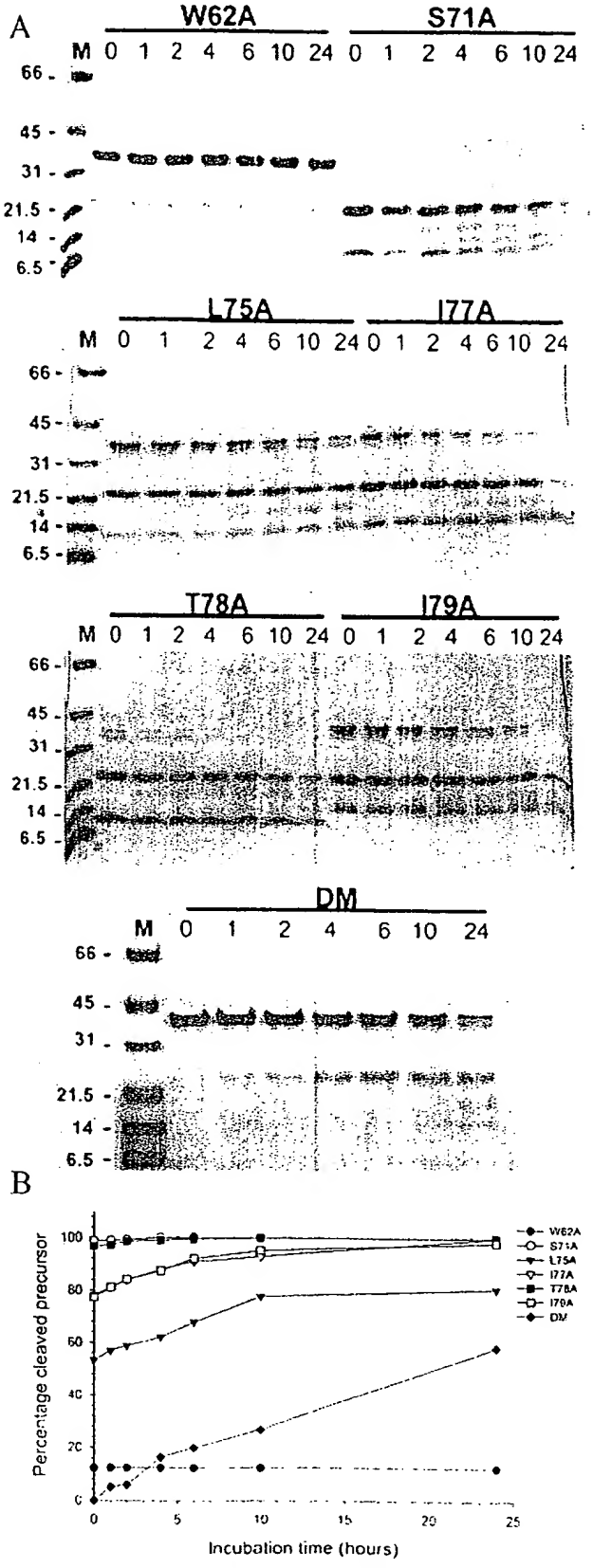


FIG. 3. Kinetics of proteolytic autoprocessing of the NS2B(H)-NS3(pro) mutant derivatives. (A) Samples of wild-type NS2B(H)-NS3pro and the mutant proteins were refolded by successive dialysis, incubated at 37°C, and analyzed on Coomassie blue-stained Tricine-SDS-PAGE gels as described in the text. Lane M, protein molecular size markers. Numbers above the lanes indicate incubation times ranging from 0 to 24 h. (B) Band intensities of wild-type NS2B(H)-NS3pro and the mutant polypeptides as observed on Tricine-SDS-PAGE gels were quantitated by densitometric analysis with the ONE-D scan program, and the fraction of cleaved NS2B(H)-NS3(pro) precursor was plotted as a function of incubation time.

TABLE 1. Kinetic parameters of the NS2B(H)-NS3pro mutant proteases[a]

Construct	Activity (% of wild-type activity)	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
Wild type	100	146 ± 5.4	1.2 ± 0.07	137 ± 3.0
S135A	ND	ND	ND	ND
W62A	ND	ND	ND	ND
S71A	118	133 ± 7.0	1.3 ± 0.03	163 ± 10.0
L75A	3.4	180 ± 4.9	0.05 ± 0.007	4.7 ± 0.5
I77A	8.4	171 ± 15.3	0.12 ± 0.01	11.5 ± 0.13
T78A	65	225 ± 5.4	1.2 ± 0.09	89.0 ± 9.0
I79A	1.9	181 ± 8.0	0.03 ± 0.004	2.6 ± 0.2
L75A/I79A	ND	ND	ND	ND

[a] Protease activity was assayed in 0.1 M Tris-HCl (pH 9.0) for 60 min at 37°C with the fluorogenic peptide GRR-AMC at concentrations ranging from 10 to 250 μM. Standard reactions contained protease at a concentration of 0.15 μM. No activity was observed with the W62A mutant and the L75A/I79A double mutant at a 1.5 μM enzyme concentration. The activity of the wild-type enzyme NS2B(H)-NS3pro with GRR-AMC was taken as 100%. ND, not detectable.

would generate products of the observed sizes. Whether the additional cleavage products are formed by internal cleavage at these sites remains to be investigated. For the L75A mutant, approximately 75% cleaved precursor was observed at 24 h of incubation. The L75A/I79A double mutant showed weak cleavage activity and yielded only 50% precursor cleavage after 24 h. The W62A mutant did not display a significant increase in the amount of autoprocessing products at 24 h. Therefore, it is likely that the W62A substitution within the NS2B cofactor results in a significant inactivation of the NS2B-NS3 protease.

Reactivity with small substrate peptides. The NS3 protease reacts with small model substrates for serine proteases in the absence of the NS2B cofactor, but cleavage efficiency of the protease towards synthetic tripeptide substrates is significantly stimulated in the presence of the 40-residue NS2B activation sequence (50). Cleavage at the NS2B/NS3 site is not a prerequisite for the reaction with small substrates, as shown by the activity of the NS2B-NS3pro precursor with 12-mer peptides (29). This protein did not display significant levels of autocleavage.

Comparison of activities with the synthetic peptide GRR-AMC between wild-type NS2B(H)-NS3pro and the mutant derivatives of NS2B revealed that the alanine substitutions affected the rate of substrate hydrolysis. Data for the kinetic parameters are presented in Table 1. The alanine substitutions mainly affected the k_{cat} values of the recombinant proteases, whereas Michaelis-Menten equilibrium constants appeared to be less affected (Fig. 4). For the T78A mutant, only a 1.5-fold increase in K_m was observed, whereas the largest changes in k_{cat} occurred in the L75A and I79A mutants, which had k_{cat}

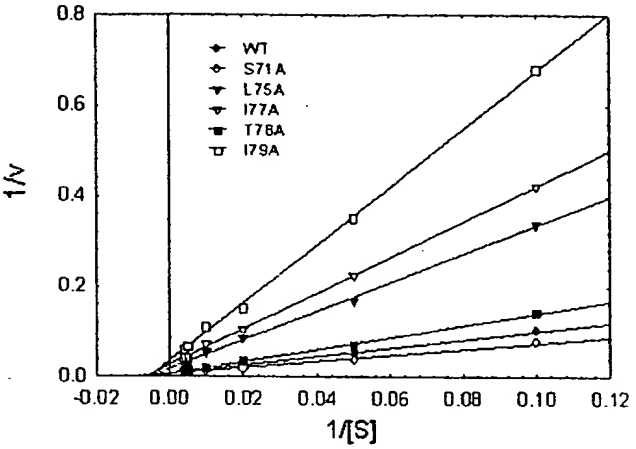


FIG. 4. Steady-state cleavage kinetics of the GRR-AMC substrate by NS2B(H)-NS3pro derivatives in vitro. Reactions were performed in triplicate at a 0.15 μM protein concentration over a range of substrate concentrations from 10 to 250 μM. Assays were performed with 60-min incubation periods, and kinetic parameters were obtained by nonlinear regression analysis of initial velocities prior to 10% substrate depletion. The L75A/I79A double mutant and the W62A mutant had no measurable activity at a 10-fold-higher enzyme concentration (1.5 μM). Data were plotted in a double reciprocal format.

values that were 24- and 40-fold lower, respectively, than that of the wild type. In the case of the S71A mutant, the replacement resulted in a slightly more active enzyme, with a k_{cat}/K_m value that was 1.2-fold higher than that of the wild type. Catalytic efficiencies expressed as k_{cat}/K_m values were substantially reduced for the L75A, I77A, and the I79A mutants, which had 30-, 12-, and 52-fold-lower efficiencies, respectively, compared to the wild type. The activity of the L75A/I79A double mutant and the W62A mutant with GRR-AMC was negligible under the conditions of the assay, and a 10-fold increase in enzyme concentration did not result in detectable conversion of the substrate.

It is unlikely that the differences in catalytic efficiency which we observed between the wild-type and mutant NS2B(H)-NS3pro proteins result simply from a distortion of the NS2B/NS3 cleavage site, since the mutations seem to affect autocleavage and reactivity with small peptides as well. Since the gel electrophoresis assay does not allow detection of *trans* cleavage activity, we cannot exclude the possibility that some of the proteolysis products shown in Fig. 3 are generated by *trans* cleavage of the NS2B(H)-NS3pro precursor. Autoprocessing at the NS2B/NS3 site is not strictly required for *trans* cleavage activity, and for mutants displaying intermediate levels of autoproteolysis, the unprocessed precursor NS2B(H)-NS3pro may also contribute to enzymatic conversion of the synthetic substrate (29). However, mutations in the NS2B/NS3 cleavage site sequence which abolish autoprocessing result in catalytically poor NS3 proteases with inefficient *trans* cleavage of less than 10% of the wild-type activity (R. Khumthong, unpublished data).

In summary, the alanine substitutions in the cofactor had greater effects on the reaction rates of the NS3 protease than on substrate binding. This would imply a model for the activation of the NS3 protease in which the cofactor contributes mainly to an arrangement of the residues of the catalytic triad that is optimized for proton transfer during substrate cleavage.

DISCUSSION

Previous studies have shown that the activity of the dengue virus NS3 serine protease depends critically on the presence of the small NS2B cofactor protein (32, 50). In this report, we further investigated the structural determinants for the interaction of the NS3 protease with the NS2B cofactor by generating alanine substitutions at selected positions possibly involved in association of the protease-cofactor complex. Effects on the enzymatic activity of the NS3 protease were determined by analysis of autoprocessing at the NS2B/NS3 site and by reaction with the synthetic peptide GRR-AMC. The enzymatic data which we obtained for both types of reactions indicate an extreme sensitivity of NS3 cleavage activity to the correct conformation of the NS2B cofactor.

First, we examined the functional relevance of the $\Phi x_3 \Phi$ motif for cofactor-induced activation of the NS3 protease. Based on structural and mutational evidence obtained for the GB virus and hepatitis C virus NS3 proteases, a consensus sequence element involved in flaviviral protease activation was recently identified (10). By comparison between the structures of the hepatitis C virus NS3 protease and the NS3/NS4A cocomplex, it was hypothesized that the binding pocket for the first Φ residue (Leu75 in dengue virus type 2) undergoes a substantial conformational change, whereas the pocket for the second Φ residue (Ile79) remains largely unchanged upon complexation with the cofactor (30, 36). The second hydrophobic amino acid was proposed to occupy the hydrophobic pocket between the two β-barrel subdomains and to contribute to stabilization of the relative orientation of these subdomains. Our data show that the alanine substitution at Ile79 had a greater effect on NS3 autocleavage activity than the substitution at Leu75, whereas the catalytic efficiency was approximately 1.8-fold lower in the I79A mutant. A drastic effect on protease activation was observed with the L75A/I79A double mutant, in which autocleavage was almost completely eliminated and enzymatic activity with the GRR-AMC peptide was not detectable under the conditions of our assay.

In addition, we examined two noncritical residues, Ile77 and Thr78, which are located within the $\Phi x_{3\Phi}$ motif. Although these two mutations had only marginal effects on autocleavage, the alanine substitution at Ile77 resulted in reduced enzyme activity, as demonstrated by a 12-fold-lower k_{cat}/K_m value compared to the wild-type enzyme. The T78A substitution yielded a less active enzyme with a 1.5-fold higher K_m value for the GRR-AMC substrate than the wild type. These results support a role for the $\Phi x_{3\Phi}$ motif in the activation mechanism of the dengue virus NS2B cofactor in which the unspecified residues also contribute to the interactions which are necessary for protease activation, a feature which discriminates the dengue virus protease from the recently analyzed GB virus NS3 protease (10). It appears that the association between the cofactor and the protease is mainly directed by hydrophobic interactions and that the mutations introduced in this region had greater effects on the catalytic efficiency of the protease than on the K_m values, suggesting that perturbation of the hydro-

phobic interactions in the Φx3Φ motif may primarily affect the geometry of the catalytic triad.

Intermediate effects on enzymatic activity, as observed for the I77A, I79A, and T78A mutants, may reflect a role of these residues not only in the conformational activation of the catalytic apparatus of the enzyme but also for the stabilization of a ternary substrate-cofactor-protease complex. For the hepatitis C virus protease, a synergistic cooperation between the cofactor and the substrate was proposed to form an induced-fit protease with optimal catalytic activity and high specificity for the polyprotein substrate (5). Therefore, in the dengue virus NS2B/NS3 protease, additional residues located outside the Φx3Φ motif may contribute to the structural rearrangements induced by cofactor binding.

Substitution of Trp62 with alanine had the strongest effect on the activity of the NS3 protease. This residue is located outside the proposed short activation sequence which is homologous to the hepatitis C virus NS4A peptide (8). A deletion construct lacking the Trp62 residue was previously shown to be inactive in cleavage assays, and the sequence comprising residues 58 to 62 was implicated in the conformational stabilization of the protease (19). The finding that Trp62 is of paramount importance for the activation mechanism also provides an explanation for the inability of the peptide Gly^{69}-Glu^{81} to replace the NS2B core sequence in vitro (32); however, it does not exclude the possibility that additional interactions at the N terminus of the NS2B(H) sequence are necessary for optimal activity, as was shown earlier for the yellow fever virus NS3 protease (17).

With the exception of the Trp62 substitution, the mutations which we introduced in the NS2B sequence of dengue virus type 2 do not simply abolish the enzymatic activity of the NS3 protease. Instead, we observed slow kinetics for autoprocessing at the NS2B/NS3 site upon prolonged incubation of the mutant enzymes. The kinetic analysis of GRR-AMC cleavage and the self-cleavage reaction indicates the existence of an inefficient catalytic machinery for both types of substrate conversion in the NS2B mutants.

In the absence of a crystallographic structure for the dengue virus NS2B-NS3 complex, we generated a model based on homology with the hepatitis C virus NS4A peptide (49). In analogy to the structure of the hepatitis C virus NS3/NS4A complex, the model predicts a threading of the cofactor in an extended conformation on a large and mainly hydrophobic surface groove of the NS3 protease formed by the N- and C-terminal domains (Fig. 5). Both hydrophobic residues Leu75 and Ile79 of the $\Phi x_3\Phi$ motif occupy a hydrophobic pocket at the domain interface. The interactions observed in this model also predict a crucial role for the Gln96 residue in the NS3 sequence by formation of a hydrogen bond to Ile79 in the NS2B main chain, an observation which would be consistent with the weak protease activity of a recently described V95A/Q96A NS3 mutant (37).

In the model presented here, the critical residue Trp62 is located in close proximity to an N-terminal cluster of proline residues, Pro10, -11, and -12, in the NS3 sequence, and the NS2B peptide is attached to a surface-exposed structure of the NS3 protease. This interaction, which is reminiscent of the N-terminal clamping observed with the hepatitis C virus NS4A peptide, may govern the correct association of the cofactor

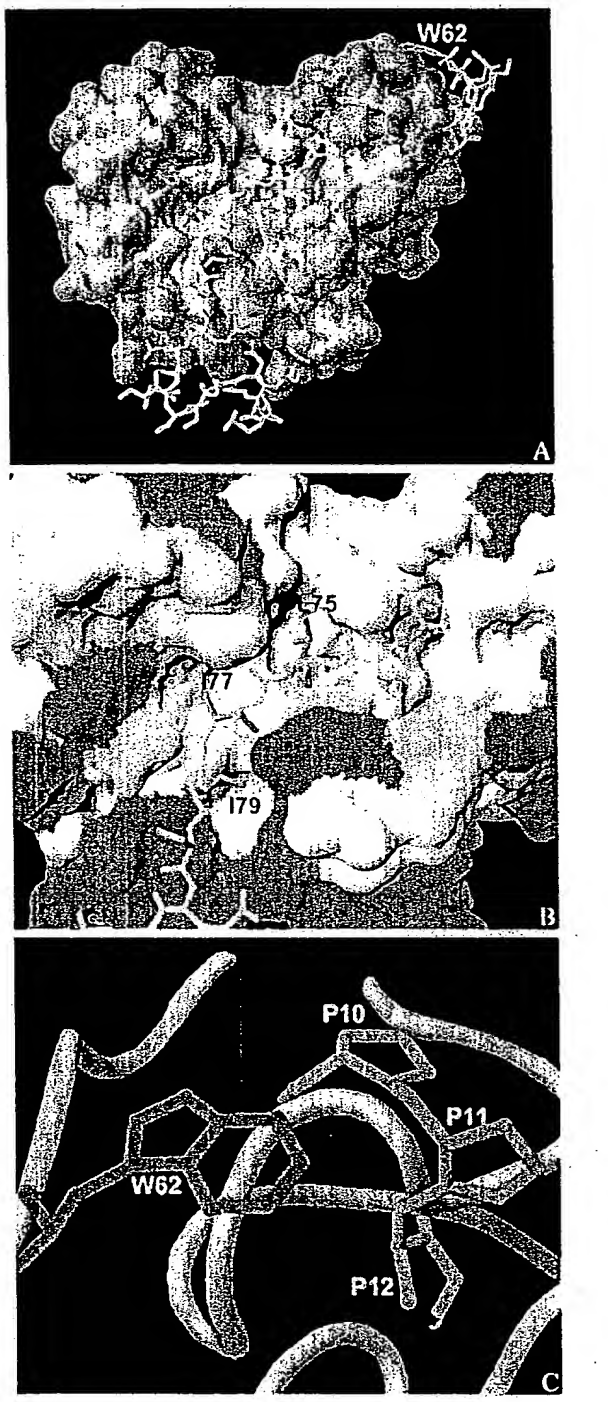


FIG. 5. Molecular model of the interaction between the NS2B core segment and the NS3 protease. The model was generated by Deepview Swiss-Pdbviewer. (A) NS2B(H) (residues 56 to 93) mapped onto the X-ray crystal structure of NS3. The surface of NS3pro is shown in blue, and NS2B(H) is shown as a ribbon in yellow. The location of Trp62 at the N terminus of NS2B(H) is shown. (B) Interaction of residues Leu75, Ile77, and Ile79 with the hydrophobic pocket of NS3. The surface of NS3 is colored by the residue type (yellow, polar; blue, basic; red, acidic; white, nonpolar); the NS2B segment is shown as a yellow stick with the side chains of Leu75, Ile77, and Ile79 in grey. (C) Association of the Trp62 residue (red) of NS2B(H) (yellow ribbon) with the N-terminal proline cluster of NS3 (side chains in green and NS3 in blue).

with the NS3 surface (49). Alternatively, it is conceivable that the binding of the cofactor contributes to a structural organization of the N-terminal region of the NS3 protein, since the N-terminal residues of NS3 display a high degree of conformational flexibility (31). For the hepatitis C virus NS3 protease, folding of the N-terminal 28 residues into a β-strand and an α-helix was observed as a result of NS4A binding (49).

Further predictions based on this model, especially changes in the geometry of the active site associated with NS2B binding, would be too speculative at the moment. Elucidation of the precise mechanism of cofactor-dependent activation of the dengue virus NS3 protease has to await the resolution of the three-dimensional structure of the NS2B-NS3 cocomplex.

The cofactor-induced activation process is comparatively well characterized for the hepatitis C virus NS3 protease by a number of structural and spectroscopic studies. Nuclear magnetic resonance experiments revealed large nonlocal structural changes leading to a catalytic triad which is better ordered in the presence of NS4A (38). Solution structures obtained with a covalently bound ketoacid inhibitor disclosed a hitherto unrecognized role for the substrate in stabilization of the catalytic machinery by the formation of hydrogen bonds within the S′ subsite of the enzyme (2). According to this model, complexation of the protease with the NS4A cofactor leads to indirect activation of NS3 and induces a conformation which is preorganized for substrate binding (5). It remains to be investigated whether similar enzyme-substrate interactions and induced-fit mechanisms contribute to active-site stabilization of the dengue virus NS2B-NS3 protease.

Taken together, our results support a model for the activation of the dengue virus protease by the NS2B cofactor which depends critically on the presence of specific residues in the cofactor core sequence rather than on the overall conformation. The residues located within the structural $\Phi x_3 \Phi$ motif play an important role in this activation process, an observation which confirms earlier findings for related flaviviral enzymes. In addition, we have shown that a single residue, Trp62, located in the N-terminal region of the NS2B core segment is of high relevance for conformational activation. The structural reasons for this unusual requirement are not entirely clear at the moment, and further studies are required to investigate the complex structure-activity relationships of the dengue virus two-component protease. These investigations are not only useful for the understanding of the cofactor-induced activation of proteolytic enzymes, but may also facilitate the development of inhibitors which interfere with the formation of viral protease complexes.

ACKNOWLEDGMENTS

We thank P. Wilairat for helpful discussions, A. Ketterman for critical review of the manuscript, C. Maharat, Prince of Songkhla University, Thailand, for N-terminal amino acid sequencing, and A. Nirachanon for excellent secretarial assistance.

This work was supported by a Senior Researcher Fellowship (to S.P.) from the Thailand Research Fund and Basic Science Grant BRG4680006 (to G.K.) from the Thailand Research Fund. Royal Golden Jubilee Ph.D. research scholarships from the Thailand Research Fund to P.N., P.W., and S.C. are gratefully acknowledged.

REFERENCES

1. **Arias, C. F., F. Preugschat, and J. H. Strauss.** 1993. Dengue virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. Virology 193:888–899.
2. **Barbato, G., D. O. Cicero, F. Cordier, F. Narjes, B. Gerlach, S. Sambucini, S. Grzesiek, V. G. Matassa, R. De Francesco, and R. Bazzo.** 2000. Inhibitor binding induces active site stabilization of the HCV protein serine protease domain. EMBO J. 19:1105–1206.
3. **Barbato, G., D. O. Cicero, M. C. Nardi, C. Steinkuehler, R. Cortese, R. De Francesco, and R. Bazzo.** 1999. The solution structure of the N-terminal proteinase domain of the hepatitis C virus (HCV) NS3 protein provides new insights into its activation and catalytic mechanism. J. Mol. Biol. 289:371–384.
4. **Bartenschlager, R., V. Lohman, T. Wilkinson, and J. O. Koch.** 1995. Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. J. Virol. 69:7519–7528.
5. **Bianchi, E., S. Orru, F. Dal Piaz, R. Ingenito, A. Casbarra, G. Biasiol, U. Koch, P. Pucci, and A. Pessi.** 1999. Conformational changes in human hepatitis C virus NS3 protease upon binding of product-based inhibitors. Biochemistry 38:13844–13852.
6. **Bianchi, E., and A. Pessi.** 2002. Inhibiting viral proteases: challenges and opportunities. Biopolymers 66:101–114.
7. **Billoir, F., R. De Chesse, H. Tolou, P. De Micco, E. A. Gould, and X. De Lamballerie.** 2000. Phylogeny of the genus flavivirus using complete coding sequences of arthropod-borne viruses and viruses with no known vector. J. Gen. Virol. 81:781–790.
8. **Brinkworth, R. I., D. P. Fairlie, D. Leung, and P. R. Young.** 1999. Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. J. Gen. Virol. 80:1167–1177.
9. **Butkiewicz, N., M. Wendel, R. Zhang, R. Jubin, J. Pichardo, E. B. Smith, A. M. Hart, R. Ingram, J. Durkin, P. Mui, M. G. Murray, L. Ramanathan, and B. Dasmahapatra.** 1996. Enhancement of hepatitis C virus NS3 proteinase activity by association with NS4A-specific synthetic peptides: identification of sequence and critical residues of NS4A for the cofactor activity. Virology 225:328–338.
10. **Butkiewicz, N., N. Yao, W. Zhong, J. Wright-Minogue, P. Ingravallo, R. Zhang, J. Durkin, D. N. Standring, B. M. Baroudy, D. V. Sangar, S. M. Lemon, J. Y. Lau, and Z. Hong.** 2000. Virus-specific cofactor requirement and chimeric hepatitis c virus/GB virus B nonstructural protein 3. J. Virol. 74:4291–4301.
11. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression and replication. Annu. Rev. Microbiol. 44:649–688.
12. **Chambers, T. J., R. C. Weir, A. Grakoui, D. W. McCourt, J. F. Bazan, R. J. Fletterick, and C. M. Rice.** 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA 87:8898–8902.
13. **Chambers, T. J., A. Grakoui, and C. M. Rice.** 1991. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J. Virol. 65: 6042–6050.
14. **Chambers, T. J., A. Nestorowitz, S. M. Amberg, and C. M. Rice.** 1993. Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. J. Virol. 67:6797–6807.
15. **Champreda, V., R. Khumthong, B. Subsin, C. Angsuthanasombat, S. Panyim, and G. Katzenmeier.** 2000. The two-component protease NS2B-NS3 of dengue virus type 2: cloning, expression and Escherichia coli and purification of the NS2B, NS3(pro) and NS2B-NS3 proteins. J. Biochem. Mol. Biol. 33:294–299.
16. **Clum, S., K. E. Ebner, and R. Padmanabhan.** 1997. Cotranslational membrane insertion of the serine protease precursor NS2B-NS3(pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. J. Biol. Chem. 272:30715–30723.
17. **Droll, D. A., H. M. Murthy, and T. J. Chambers.** 2000. Yellow fever virus NS2B-NS3 protease: charged-to-alanine mutagenesis and deletion analysis define regions important for protease complex formation and function. Virology 275:335–347.
18. **Falgout, B., M. Pethel, Y. M. Zhang, and C. J. Lai.** 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65:2467–2475.
19. **Falgout, B., R. H. Miller, and C. J. Lai.** 1993. Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. J. Virol. 67:2034–2042.
20. **Gorbalenya, A. E., A. P. Donchenko, E. V. Koonin, and V. Blinov.** 1989. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. Nucleic Acids Res. 17:3889–3897.
21. **Gubler, D. J.** 2002. Epidemic dengue/dengue haemorrhagic fever as a public

health, social and economic problem in the 21st century. Trends Microbiol. 10:100–103.
22. Halstead, S. B. 1988. Pathogenesis of dengue: challenges to molecular biology. Science 239:476–481.
23. Halstead, S. B. 1997. Epidemiology of dengue and dengue hemorrhagic fever, p. 23–44. *In* D. J. Gubler and G. Kuno (ed.), Dengue and dengue hemorrhagic fever. CAB International, New York, N.Y.
24. Ingallinella, P., S. Altamura, E. Bianchi, M. Taliani, R. Ingenito, R. Cortese, R. De Francesco, C. Steinkuehler, and A. Pessi. 1998. Potent peptide inhibitors of the human hepatitis C virus NS3 protease are obtained by optimizing the cleavage products. Biochemistry 37:8906–8914.
25. Ingallinella, P., D. Fattori, S. Altamura, C. Steinkuehler, U. Koch, D. Cicero, R. Bazzo, R. Cortese, E. Bianchi, and A. Pessi. 2002. Prime site binding inhibitors of a serine protease: NS3/4A of hepatitis C virus. Biochemistry 41:5483–5492.
26. Irie, K., P. M. Mohan, Y. Sasaguri, R. Putnak, and R. Padmanabhan. 1989. Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). Gene 75:197–211.
27. Kadare, G., and A. L. Haenni. 1997. Virus-encoded RNA helicases. J. Virol. 71:2583–2590.
28. Khumthong, R., C. Angsuthanasombat, S. Panyim, and G. Katzenmeier. 2002. In vitro determination of dengue virus type 2 NS2B-NS3 protease activity with fluorescent peptide substrates. J. Biochem. Mol. Biol. 35:206–212.
29. Khumthong, R., P. Niyomrattanakit, S. Chanprapaph, C. Angsuthanasombat, S. Panyim, and G. Katzenmeier. 2003. Steady-state cleavage kinetics for dengue virus type 2 NS2B-NS3(pro) serine protease with synthetic peptides. Protein Peptide Lett. 10:19–26.
30. Kim, J. L., K. A. Morgenstern, C. Lin, T. Fox, M. D. Dwyer, J. A. Landro, S. P. Chambers, W. Markland, C. A. Lepre, E. T. O'Malley, S. L. Harbeson, C. M. Rice, M. A. Murcko, P. R. Caron, and J. A. Thomson. 1996. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell 87:343–355.
31. Krishna Murthy, H. M., S. Clum, and R. Padmanabhan. 1999. Dengue virus NS3 serine protease: crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects. J. Biol. Chem. 274:5573–5580.
32. Leung, D., K. Schroeder, H. White, N. X. Fang, M. J. Stoermer, G. Abbenante, J. L. Martin, P. R. Young, and D. P. Fairlie. 2001. Activity of recombinant dengue virus NS3 protease in the presence of a truncated NS2B cofactor, small peptide substrates and inhibitors. J. Biol. Chem. 276:45762–45771.
33. Li, H., S. Clum, S. You, K. E. Ebner, and R. Padmanabhan. 1999. The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. J. Virol. 73:3108–3116.
34. Lin, C., S. M. Amberg, T. J. Chambers, and C. M. Rice. 1993. Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. J. Virol. 67:2327–2335.
35. Lobigs, M. 1993. Flavivirus premembrane protein cleavage and spike heterodimer secretion require the function of the viral proteinase NS3. Proc. Natl. Acad. Sci. USA 90:6218–6222.
36. Love, R. A., H. E. Parge, J. A. Wickersham, Z. Hostomsky, N. Habuka, E. W. Moomaw, T. Adachi, and Z. Hostomska. 1996. The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. Cell 87:331–342.
37. Matusan, A. E., P. G. Kelley, M. J. Pryor, J. C. Whisstock, A. D. Davidson, and P. J. Wright. 2001. Mutagenesis of the dengue virus type 2 NS3 proteinase and the production of growth-restricted virus. J. Gen. Virol. 82:1647–1656.
38. McCoy, M. A., M. M. Senior, J. J. Gesell, L. Ramanathan, and D. F. Wyss. 2001. Solution structure and dynamics of the single-chain hepatitis C virus NS3 protease NS4A cofactor complex. J. Mol. Biol. 305:1099–1110.
39. Pessi, A. 2001. A personal account of the role of peptide research in drug discovery: the case of hepatitis C. J. Peptide Sci. 7:2–14.
40. Preugschat, F., C. W. Yao, and J. H. Strauss. 1990. In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J. Virol. 64:4364–4374.
41. Ryan, M. D., S. Monaghan, and M. Flint. 1998. Virus-encoded proteinases of the Flaviviridae. J. Gen. Virol. 79:947–959.
42. Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234:779–815.
43. Schaegger, H., and G. von Jagow. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. 166:368–379.
44. Shimizu, Y., K. Yamaji, Y. Masuho, T. Yokota, H. Inoue, K. Sudo, S. Satoh, and K. Shimotohno. 1996. Identification of the sequence on NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 protease. J. Virol. 70:127–132.
45. Steinkuehler, C., A. Urbani, L. Tomei, G. Biasiol, M. Sardana, E. Bianchi, A. Pessi, and R. De Francesco. 1996. Activity of purified hepatitis C virus protease NS3 on peptide substrates. J. Virol. 70:6694–6700.
46. Teo, K. F., and P. J. Wright. 1997. Internal proteolysis of the NS3 protein specified by dengue virus 2. J. Gen. Virol. 78:337–341.
47. Valle, R. P. C., and B. Falgout. 1998. Mutagenesis of the NS3 protease of dengue virus type 2. J. Virol. 72:624–632.
48. Wengler, G., G. Czaya, P. M. Farber, and J. H. Hegemann. 1991. In vitro synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. J. Gen. Virol. 72:851–858.
49. Yan, Y., Y. Li, S. Munshi, V. Sardana, J. L. Cole, M. Sardana, C. Steinkuehler, L. Tomei, R. De Francesco, L. C. Kuo, and Z. Chen. 1998. Complex of NS3 protease and NS4a peptide of BK strain hepatitis C virus: a 2.2 A resolution structure in a hexagonal crystal form. Protein Sci. 7:837–847.
50. Yusof, R., S. Clum, M. Wetzel, H. M. Murthy, and R. Padmanabhan. 2000. Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids. J. Biol. Chem. 275:9963–9969.
51. Zhang, L., P. M. Mohan, and R. Padmanabhan. 1992. Processing and localization of dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. J. Virol. 66:7549–7554.

Insect Molecular Biology (2002) 11(3), 197–205

Association of midgut defensin with a novel serine protease in the blood-sucking fly *Stomoxys calcitrans*

J. V. Hamilton, R. J. L. Munks, S. M. Lehane and M. J. Lehane

School of Biological Sciences, University of Wales, Bangor UK

Abstract

Using ELISA we provide direct evidence that the midgut defensins of the blood-sucking fly *Stomoxys calcitrans* are secreted into the gut lumen. We show that midgut defensin peptide levels increase up to fortyfold in response to a blood meal but not to a sugar meal. The data suggests the midgut defensin genes are post-transcriptionally regulated and that their function is protection of the stored blood meal from bacterial attack while it awaits digestion. Using recombinant defensins produced in *Pichia pastoris* we demonstrate that while in the gut cells the midgut defensins are bound in an SDS-stable complex to proteins with an apparent molecular weight of > 26 kDa from which they are released when secreted into the gut lumen. This > 26 kDa protein (Ssp3) has been cloned and sequenced and is a member of the serine protease S1 family with homologies to multiple insect proteases and to vertebrate trypsins and elastases.

Keywords: defensin, serine protease, immunity, blood-sucking, midgut.

Introduction

The major immune responses of insects, or at least the most studied, are mounted from the insect fat body. Challenge to the insect leads to transcriptional up-regulation of a series of genes and the *de novo* synthesis and secretion of a battery of immune peptides from the fat body (Hoffmann & Reichhart, 1997). These insect immune responses are non-clonal and have many similarities to the innate immune system of vertebrates (Medzhitov & Janeway, 1997). Insects also have well developed epithelial immune responses (Brey *et al.*, 1993; Lehane *et al.*, 1997; Tzou *et al.*, 2000). In *Drosophila* all of the epithelial surfaces of the insect body produce antimicrobial peptides usually as a subset of the total number of antimicrobial peptides in the insect; the subset is usually complementary containing both anti-Gram positive and anti-Gram negative activities (Tzou *et al.*, 2000). Among these epithelia only *Drosophila* midgut expresses all of the antimicrobial peptides (Tzou *et al.*, 2000) and this reflects the relative vulnerability of the gut to infection, which is common to all metazoa. Given that the midgut is the primary site of entry for the most important insect borne parasites including malaria, trypanosomes and arbovirus for example, it is surprising how little we know about insect midgut immunity, particularly in blood-sucking insects. We do know that agglutinins play an antiparasitic role in the midgut of some insects (Maudlin, 1991), that peritrophic matrix has a defensive function (Lehane, 1997), that the anti-Gram positive enzyme lysozyme may be present and that *Plasmodium* ookinetes may, on occasion, be lysed by undefined mechanisms in the midgut (Vernick *et al.*, 1995).

In blood-sucking insects secretion of antimicrobial peptides in the midgut epithelium has been demonstrated in *Stomoxys calcitrans* (Lehane *et al.*, 1997), *Anopheles gambiae* (Dimopoulos *et al.*, 1997) and *Glossina morsitans morsitans* (Hao *et al.*, 2001). The two midgut specific defensins reported from the anterior midgut (reservoir) of the stable fly *S. calcitrans* were particularly unusual in being specific to that tissue and being constitutively produced (Lehane *et al.*, 1997; Munks *et al.*, 2001). Those studies were performed exclusively on mRNA from these genes. Consequently there is no direct proof that those defensin peptides are indeed used in the midgut lumen. Also there is mounting evidence that many genes in the midgut of blood-sucking insects are post-transcriptionally controlled so that meaningful data on gene product levels can only come from studies of the peptides themselves (Muller *et al.*, 1995; Lehane *et al.*, 1998; Noriega & Wells, 1999). In this study we set out to use protein-based techniques to obtain a clearer view of the biology of the *S. calcitrans* midgut defensins.

Received 11 October 2001; accepted after revision 18 December 2001.
Correspondence: M. J. Lehane, School of Biological Sciences, University of Wales, Bangor LL57 2UW, UK. Tel.: + 1248 382309; fax: + 1248 370731; e-mail: m.j.lehane@bangor.ac.uk

Results

Recombinant proteins

The *Pichia* system produced low levels of recombinant protein; range 0.1–16.7 μg/l for Smd1 and 0.6–6.2 μg/l for Smd2. Mass spectrophotometric analysis (data not shown) suggests both recombinant defensins were modified compared to the native forms. Recombinant Smd1 was truncated losing its first four amino acids. Recombinant Smd2 was glycosylated (internal N-linked site at amino acid positions 12–14 in the mature peptide) whereas the native form is not (Lehane *et al.*, 1997). Changing expression medium from minimal methanol to a medium enriched with 1% casamino acids made no difference to either modification. Despite these structural changes from the native defensins these recombinant proteins provided useful positive controls in the antisera work as described below.

Specificity and cross-reactivity of Smd1 and Smd2 rabbit antisera

The specificity of the antisera was determined by Western blotting (Fig. 1) and ELISA (Fig. 2), the cross-reactivity of the two sera with each other was also determined in ELISA (Fig. 2). Western blotting with the sera against recombinant Smd1 and Smd2 demonstrated that each serum recognized a single band, at approximately 4 kDa, which corresponded with the predicted molecular weights of 4736 Da for Smd1 (Fig. 1A) and 4237 Da for Smd2 (Fig. 1B) (Lehane *et al.*, 1997). These protein bands were not recognized by either the control serum or the pre-immune sera (Fig. 1C, lanes 1 & 2 and 3 & 4, respectively).

In ELISA, checkerboard titration of antisera and antigen revealed that the optimum serum dilution was 1 : 5000 when an antigen coating concentration of 5 μg/ml was used (data not shown). ELISA demonstrated that Smd1 and Smd2 antiserum recognized both the synthetic peptide against which they were raised (data not shown) and the recombinant Smd1 and Smd2 peptides (Fig. 2). Control and pre-immune sera were used to establish baseline background activity. In summary, the two sera used in this study were sensitive, specific and did not cross-react with each other (Figs 1 and 2).

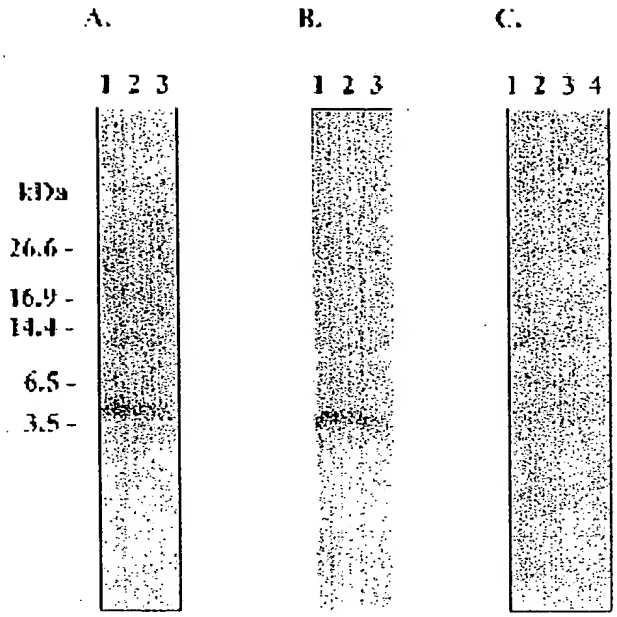


Figure 1. Specificity of Smd1 and Smd2 antisera. Western blot of 0.87 μg/lane of recombinant Smd1 (Panels A & C) or Smd2 (Panel B) peptide separated on 16.5%T, 3%C Tris-Tricine SDS-PAGE probed with A, anti-Smd1 sera (MJL3), 1 : 50 dilution (lane 1); 1 : 100 dilution (lane 2); 1 : 200 dilution (lane 3). B, anti-Smd2 sera (MJL8), 1 : 50 dilution (lane 1); 1 : 100 dilution (lane 2); 1 : 200 dilution (lane 3). C, control sera (MJL10) 1 : 50 dilution (lane 1); 1 : 100 dilution (lane 2). Pooled pre-immune sera 1 : 50 dilution (lane 3) and 1 : 100 dilution (lane 4).

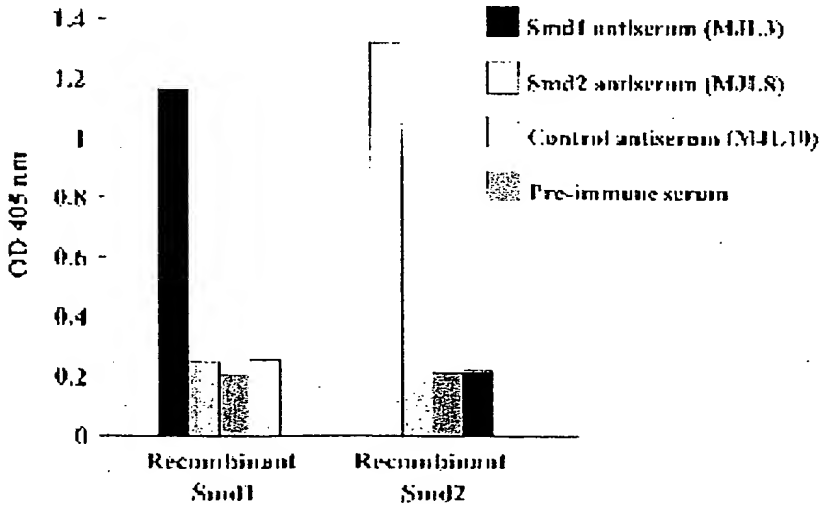


Figure 2. Sensitivity, specificity and cross-reactivity of Smd1 and Smd2 antisera in ELISA. Corrected ELISA optical densities, obtained as described in ELISA experimental procedures section. Graph 1: recombinant Smd1 peptide probed with Smd1 antiserum (MJL3, black bar), Smd2 antiserum (MJL8, white bar), control antiserum (MJL10, light grey bar) and pre-immune serum (dark grey bar). Graph 2: recombinant Smd2 peptide Smd2 antiserum (MJL8, white bar), Smd1 antiserum (MJL3, black bar), control antiserum (MJL10, light grey bar) and pre-immune serum (dark grey bar).

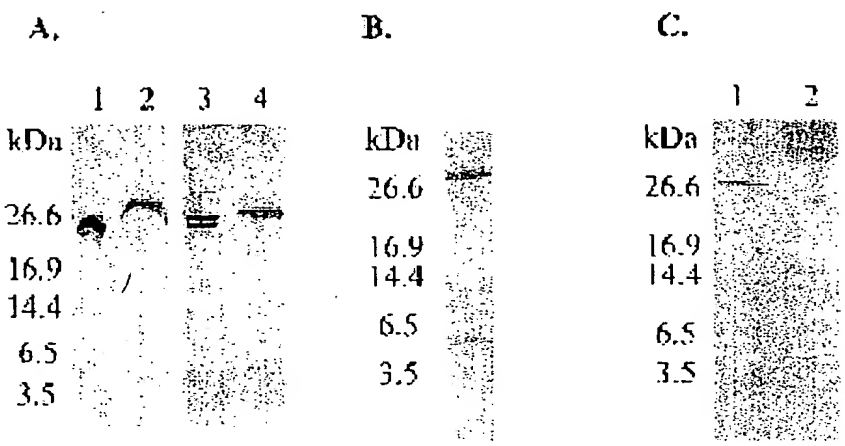


Figure 3. Colocalization of defensin and Ssp3 in the midgut of Stomoxys calcitrans. Western blots of midgut samples separated on 16.5%T, 3%C Tris-Tricine SDS-PAGE. A, effect of spiking whole reservoir zone homogenate with recombinant Smd1: 10 µg homogenate from unfed flies (lane 1); 10 µg homogenate from unfed flies spiked with 2.5 µg recombinant Smd1 peptide (lane 2); 10 µg homogenate from fed flies (24 h post blood meal) (lane 3); 10 µg homogenate from fed flies spiked with 2.5 µg recombinant Smd1 peptide (lane 4). B, presence of Smd1 defensin (≈ 4 kDa band) and Ssp3 (> 26 kDa band) in 50 µg reservoir zone homogenate from fed flies (24 h post blood meal). C, presence of Ssp3 (> 26 kDa band) in 10 µg reservoir zone tissue (lane 1) and Smd1 defensin (~4 kDa band) in 10 µg of reservoir zone contents (lane 2) from fed flies (24 h post blood meal).

SDS-stable complexes with Smd1 defensin

Western blotting of 10 µg per lane of reservoir zone homogenates revealed the presence of a > 26 kDa band(s) when probed with Smd1 antiserum (Fig. 3A, lane 1) and Smd2 antiserum (data not shown), but did not demonstrate the presence of the expected ≈ 4 kDa band (defensin). When these reservoir zone homogenates (10 µg per lane) were spiked with recombinant Smd1 peptide (2.5 µg per lane) and probed with Smd1 antiserum, the apparent molecular weights of these > 26 kDa bands increased (Fig. 3A, lanes 1 & 2). This was consistent for midgut homogenates from both unfed (Fig. 3A, lanes 1 & 2) and fed flies (Fig. 3A, lanes 3 & 4). In these 'spiking' experiments we did not observe the presence of the expected ≈ 4 kDa defensin band. When we increased the amount of midgut homogenate loaded on to the gels to 50 µg per lane and probed the blot with Smd1 antisera we detected the presence of both the > 26 kDa band and the ≈ 4 kDa defensin band (Fig. 3B).

After establishing the presence of Smd1 defensin in the midgut we repeated the experiment using flies dissected 24 h after a blood meal, but this time carefully separated the reservoir tissue from the reservoir contents. We now found that the > 26 kDa band was only present in the reservoir tissue (Fig. 3C, lane 1) whilst the 4 kDa band was only present in the gut contents (Fig. 3C, lane 2). These results coupled with the results from the 'spiking' experiment (Fig. 3A) led us to hypothesize that Smd1 peptide is able to associate in an SDS-stable complex with this higher molecular weight protein. Although no other specific defensin–protein SDS-stable complexes are described in the literature, other SDS-stable peptide-protein complexes are well known from vertebrate studies (Scott *et al.*, 1999).

The presence of a > 26 kDa doublet in some of the experiments may be due to the detection of a modified form of the > 26 kDa protein, for example glycosylation or other post-translational modifications, or may simply be due to incomplete reduction of disulphide bonds during sample preparation. It is also possible that there may be some antibody cross-reactivity, either specific cross-reactivity of a similar or related protein or non-specific cross-reactivity. However, the observation of an increase in the apparent molecular weight of both bands in the doublet after spiking with recombinant Smd1 peptide (Fig. 3A) suggests that the doublet represents a modified or partially reduced form of the > 26 kDa protein.

Western blotting revealed that Smd1 defensin was only detected in the reservoir zone homogenates and was not present in other regions of the midgut including the proventriculus, thoracic midgut, opaque zone or lipoid zone (data not shown). In addition, Smd1 defensin was not present in the fat body, malphigian tubules, brain or thoracic flight muscle (data not shown).

Purification and N-terminal amino acid sequence

The > 26 kDa protein that complexed with Smd1 was purified by repeated excision of the band from SDS-PAGE and electro-elution until a single band was revealed by SDS-PAGE. The identity of the band was confirmed by Western blotting a sub-sample with anti-Smd1 serum before obtaining the following partial N-terminal amino acid sequence by Edman sequencing: IVGGNAFAHEGQFPHQVSS. Blast searches revealed that amino acids 1–17 of the > 26 kDa protein had 76% identity with amino acids 50–66 of serine protease SP24D from *Anopheles gambiae*, suggesting that this peptide fragment was from a serine protease.

Full length clone sequence and alignment with other serine proteases

Using a degenerate sense primer designed from the least conserved portions of the Edman sequence and the M13–20 primer a full length cDNA was cloned from the library (accession number AY044834). It was 881 nucleotides long, coding for a prepro-protein of 254 amino acids. The predicted molecular weight of the full-length prepro-protein

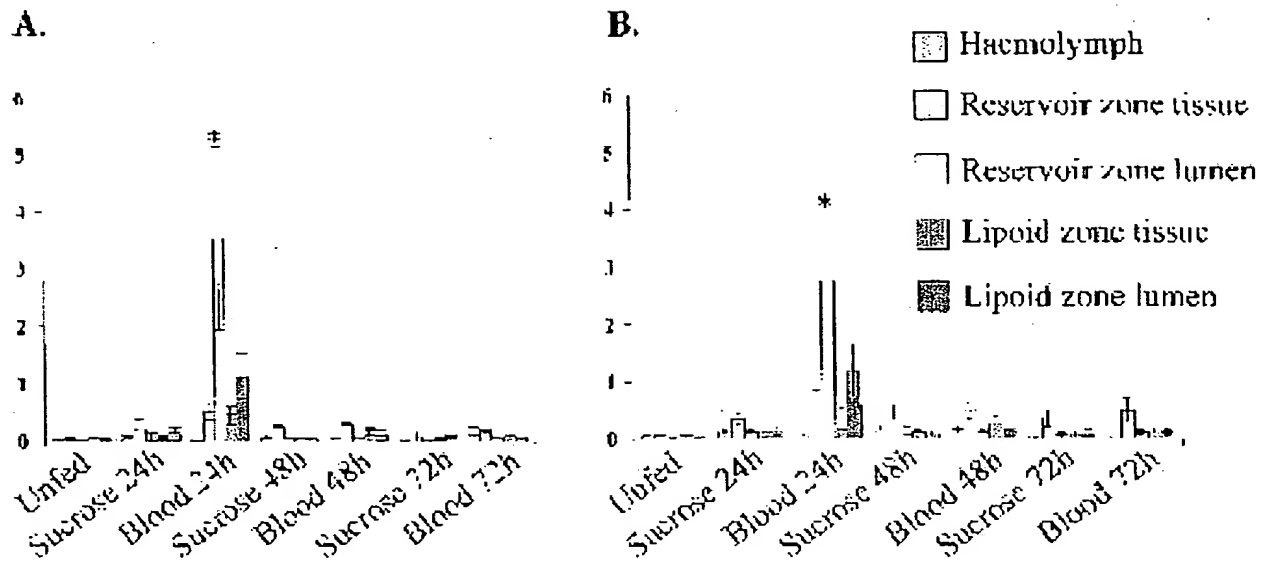


Figure 4. Mean Smd1 and Smd2 levels in adult *Stomoxys calcitrans* haemolymph, reservoir zone tissue, reservoir zone lumen, lipoid zone tissue and lipoid zone lumen. A, mean and standard errors of Smd1 levels from unfed (UF), sucrose fed (S) and blood fed (B) flies at 24, 48 and 72 h after feeding. Results are from five experiments. B, mean and standard errors of Smd2 levels.* indicates a significant difference from haemolymph, reservoir zone tissue and lipoid zone tissue ($P < 0.05$, one-way ANOVA and Fishers pair-wise comparison).

is 27 523.30 Da with a theoretical pI of 5.85 and charge of −4.25. SignalP analysis V 1.1 (Nielsen *et al.*, 1997) of the sequence suggests that the leader sequence consists of amino acid residues 1–23, which is cleaved between a serine and alanine residue. We suggest the six amino acid activation peptide ARPRPR is cleaved from the mature protein between the arginine and isoleucine at position 30. The theoretical molecular weight, charge and pI of the signal peptide would be 2601.40 Da, +0.91 and 8.83, respectively, and for the activation peptide they would be 750.90 Da, +2.91 and 12.40, respectively. The mature protein has a theoretical molecular mass of 24 205.00 Da, a charge of −8.25 and a pI of 5.00.

Blastp was used to search GenBank with the > 26 kDa sequence (Altschul *et al.*, 1997) and this revealed homology of the > 26 kDa protein with members of the serine protease S1 family (39% identity with chymotrypsin 1 and 37% identity with serine protease SP24D from *Anopheles gambiae*). Consequently we have named the protein *Stomoxys* serine protease 3 (Ssp3) (accession number AY044834).

Localization of Smd1 and Smd2 peptides and Ssp3 mRNA

The distribution of Smd1 and Smd2 in the adult fly was investigated using ELISA (Fig. 4). Low levels of Smd1 and Smd2 were present in the reservoir zone tissues in unfed and sucrose fed flies. However, in blood fed flies these levels increased at least twofold within 24 h of feeding. The highest levels of Smd1 and Smd2 were detected in the reservoir zone lumen where there was a fortyfold increase in Smd1 and a twenty-threefold increase in Smd2 24 h post blood meal. Gut lumen defensin levels did not increase in response to a sucrose meal. This suggests that Smd1 and Smd2 are up-regulated only in response to the blood meal and secreted into the midgut lumen. The finding of defensins in lipoid zone lumen presumably reflects passage down the gut lumen from the reservoir because no defensin mRNA is present in lipoid zone tissue (Lehane *et al.*, 1997; Munks *et al.*, 2001). Compared to unfed flies the levels of Smd1 and Smd2 in the reservoir tissues, but not the lumen, remain elevated up to 72 h post blood meal. RT-PCR shows that Ssp3 is not restricted to the anterior midgut but is found in all tissues tested in the adult fly (Fig. 5).

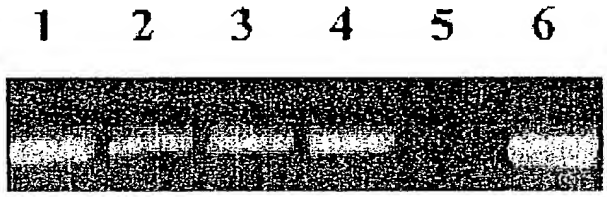


Figure 5. RT-PCR of Ssp3 mRNA. mRNA was prepared from anterior midgut (cardiac, thoracic and reservoir zones, lane 1), posterior midgut (opaque and lipoid zones, lane 2), fat body (lane 3) and carcass (lane 4). Negative control (no mRNA, lane 5) and positive control (Ssp3 clone, lane 6).

Importance of Smd1 and Smd2 in overall midgut antimicrobial response

Zone inhibition assays confirm the presence of antimicrobial activity in midgut homogenates (Fig. 6, zone A). Neither the control antibody nor isotonic saline had any effect on the growth of *M. luteus*. When midgut homogenates were incubated with anti-Smd1 antibody there was a 53% reduction

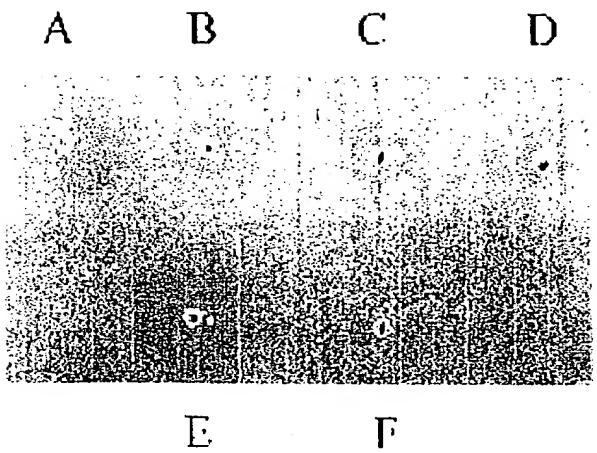


Figure 6. Zone inhibition assay using *M. luteus* as the test organism. Samples in wells were 4 midgut equivalents (w/w protein) from A, gut homogenate alone; B, gut homogenate incubated with Smd1 antibody; C, gut homogenate incubated with Smd2 antibody; D, gut homogenate incubated with both Smd1 and Smd2 antibodies; E, gut homogenate incubated with control antibody; F, Isotonic saline alone.

in the zone of inhibition (Fig. 6, zone B), when homogenates were incubated with anti-Smd2 antibody there was a 47% reduction (Fig. 6, zone C) and when the homogenates were incubated with both antibodies there was a 59% reduction in antimicrobial activity (Fig. 6, zone D) suggesting the presence of other anti-Gram positive agents in the midgut. We were unable to produce recombinant defensins that were exactly the same as the native defensins, therefore we were unable to establish that the concentration of antibodies used in these inhibition assays would be sufficient to inhibit purified defensins. Interestingly, when Smd1 and Smd2 were inhibited by the antibodies *M. luteus* appeared to encroach back into the inhibition zone suggesting that the other antimicrobial agents in the midgut may be bacteriostatic rather than bacteriocidal or that they are relatively unstable.

Discussion

Smd1 and 2 peptides are constitutively produced but ingestion of the blood meal induces up to a fortyfold increase in defensin production in the gut (Fig. 4). The data presented here combined with previous data (Lehane *et al.*, 1997; Munks *et al.*, 2001) suggests that production of these proteins may be regulated post-transcriptionally. Post-transcriptional regulation of genes appears to be a common phenomenon in the midgut of blood-sucking insects (Muller *et al.*, 1995; Lehane *et al.*, 1998; Noriega & Wells, 1999) and may be a consequence of the selective advantage to haematophagous insects of rapid blood meal digestion (Lehane, 1991).

Although we and others have previously demonstrated antimicrobial gene mRNA in insect midgut (Lehane *et al.*, 1997; Dimopoulos *et al.*, 1997; Hao *et al.*, 2001) the data presented (Figs 3C and 4) is the first direct evidence that an antimicrobial peptide is secreted into the insect gut lumen. The massive response to the blood meal contrasted with the weak response to sugar meals (Fig. 4), and the high concentration of defensins in the anterior midgut (where the undigested blood meal is stored) compared to the posterior midgut (where the blood meal is digested), supports the conclusion that midgut defensins help to protect the blood meal from bacterial attack during the 24 h before the meal is fully digested by the fly (Lehane *et al.*, 1997). This conclusion is strengthened by the fact that mRNA for Smd1 and 2 are only found in the anterior midgut (Lehane *et al.*, 1997; Munks *et al.*, 2001) and by the rapid decline in defensin proteins in the gut lumen 24 h post blood meal when the final portions of the stored blood meal have been passed through the gut for digestion. The data presented (Fig. 6) suggests that these two defensins may form only a part, 60% as crudely estimated by zone assays (Fig. 6, zone D), of the anti-Gram positive activity to be found in the anterior midgut. However, we were unable to establish that the concentrations of antibodies used in this assay would fully inhibit purified defensins, therefore these results merely lead us to speculate that there is other anti-Gram positive antimicrobial activity in the anterior midgut.

Our data, particularly the spiking experiments (Fig. 3C), suggest that while in the reservoir tissues the defensins are bound to Ssp3. The presence of a doublet in the spiking experiment (Fig. 3A) may reflect the presence of a modified form of Ssp3. SDS-stable complexes of this sort are well known (Kato *et al.*, 2001). Once the material is secreted into the midgut lumen the two proteins become dissociated (Fig. 3B), which suggests that the Ssp3 defensin aggregate is not the active unit in the midgut lumen. The discrepancy between the apparent molecular weight of Ssp3 at > 26 kDa and the predicted molecular weight of 24.205 kDa may be due to the presence of bound defensin, possible post-translational modifications of the protease, incomplete reduction of the protease during sample preparation or any combination of these possibilities. At present we do not know the mechanism of the interaction(s) between Smd1 defensin peptide and the Ssp3 serine protease. Indeed the number of binding sites available to the defensin on the serine protease and whether the defensin can bind to itself when associated with the protease remains to be elucidated.

The full sequence of mature Ssp3 was used to search GenBank using Blastp (Altschul *et al.*, 1997), which produced multiple alignments with serine proteases, notably insect and vertebrate trypsins and vertebrate elastases. These included interesting examples such as Met-ase-1 (granzyme M) a serine protease from cytolytic granules of rat CD3(−) large granular lymphocytes (Kelly *et al.*, 1996) believed to play a part in innate immune responses in

vertebrates (Sayers *et al.*, 2001). Inspection of these aligned sequences shows Ssp3 contains the catalytic triad, His, Asp and Ser in that order and that the highly conserved regions surrounding the His and Ser residues (which are typical of serine proteases) are also conserved (Kraut, 1977). Ssp3 also has the six highly conserved cysteine residues at the positions that would allow the formation of the three cysteine bonds typical of invertebrate serine proteases and differentiating them from the vertebrate enzymes, which have four such bonds. We conclude that Ssp3 belongs to the peptidase family S1; its particular substrate specificity and function needs to be determined empirically. We note that blood meal digestion does not occur in the reservoir region of the fly where Ssp3 is produced suggesting a non-digestive function for Ssp3. This is supported by both the widespread distribution of Ssp3 mRNA in the body (Fig. 5) and by the relatively poor homology of this serine protease to the two already described digestive serine proteases from *S. calcitrans* (Lehane *et al.*, 1998). It is unlikely that Ssp3 is involved in haemolysis of the blood meal in the reservoir because haemolysis does not occur in this zone (M. J. Lehane, unpublished observation). We have considered the possibility that Ssp3 may play a role in the enzymatic activation of defensin from the pro- to the mature form, which is known to be a key regulatory step in vertebrates (Wilson *et al.*, 1999). It is possible that Ssp3 may be involved in similar activity in the fat body as it is also present in this tissue. Interestingly the cleavage site leading to mature Smd1 is Ala-Ala, which is the preferred site of elastase and the homology searches suggest good homology between Ssp3 and vertebrate elastases. This may be a function of Ssp3. However, enzymatic activation would not require an SDS-stable linkage between the two molecules. A possible function of this tight association may be the inactivation of either the defensin or the trypsin or both while they are within the tissues. Although we are not aware of any other examples of defensin–serine protease associations other SDS-stable complexes have been extensively described for vertebrates. In particular, the interactions of serpins with proteases such as the inhibition of granulocyte proteases by the intracellular serpin, proteinase inhibitor 6 (PI-6) (Scott *et al.*, 1999).

Serine protease has already been immunocytochemically located in *S. calcitrans* reservoir zone secretory granules (Jordao *et al.*, 1996) and it seems probable that the Ssp3 defensin aggregate is localized there. Colocalization of proteases with antibacterial peptides within single secretory vesicles is well documented in vertebrates. For example, it has been shown in mouse Paneth cell granules that defensins (cryptidins) require proteolytic activation by the metalloproteinase matrilysin, which is colocalized in its granules (Wilson *et al.*, 1999). An example of a different type of association is given by the azurophil granule, a specialized lysosome of neutrophils. It contains two families of antimicrobial proteins, each with four members. The defensins, comprising human neutrophil protein 1, –2, –3 and –4, on the one hand and the serprocidins, comprising cathepsin G, elastase, proteinase 3 and azurocidin, on the other (Gabay & Almeida, 1993). Interestingly antibacterial activity has been reported in an antimicrobial peptide-associated serine protease from the midgut cells of the fly *Sarcophaga peregrina* (Tsuji *et al.*, 1998). The activity of this molecule is an intrinsic characteristic of the protein not related to its protease activity. This protease is found in the yellow body, which is formed from primordial adult midgut cells in the puparium. In a fascinating parallel with our studies it has been found that an antibody to the antibacterial peptide sarcotoxin IA (cecropin family) binds to this 26 kDa protease (Nakajima *et al.*, 1997). The authors suggest either cross reactivity of their antibody with the protease, or that sarcotoxin 1 A forms an SDS-stable complex to the 26 kDa protease. Our data strongly supports the latter hypothesis for the following reason. Smd1 is limited to the reservoir region (Lehane *et al.*, 1997; Munks *et al.*, 2001) while Ssp3 is widely distributed in the tissues of the adult (Fig. 5) and Western analysis shows banding only in the reservoir region where both are present (Fig. 3B). So evid ence is accumulating in insects that antibacterial peptides and proteases are colocalized in tissues and, as in vertebrates, that the associated protease may also be antibacterial.

Interactions of insect immune molecules with proteases deserves further study, particularly if immune systems are to be targeted for genetic manipulation in vector-borne disease control.

Experimental procedures

Insects

S. calcitrans was cultured as previously described (Blakemore *et al.*, 1993). The artificial blood meal (Lehane *et al.*, 1998) and sugar meals were made with high purity water (18MΩ).

Production of recombinant defensins

Recombinant proteins were produced using a commercial *Pichia pastoris* system (InVitrogen). The full Smd1 or Smd2 sequence preceded by the sequence defining the *KEX2* cleavable segment (Glu-Lys-Arg) of the α-factor mating signal was generated by PCR and inserted into the *Xho*I/SnaB1 or the *Xho*I/*Eco*RI site, respectively, of the plasmid pPIC9. pPIC9 was linearized with *Sal*I and transformed into GS115. *Pichia* was grown in minimum glycerol medium and the inserted gene expressed in minimal methanol medium. The product was purified by HPLC (Lehane *et al.*, 1997) and protein expression levels were determined by the Bradford method (Bio-Rad).

Antibody production

Unique regions of Smd1 and Smd2 were identified by amino acid sequence comparisons and short amino acid sequences (10mers) were commercially synthesized (MWG-Biotech, Germany). The

synthetic peptides were conjugated to bovine thyroglobulin (Sigma-Aldrich, UK) via gluteraldehyde (Adrian, 1997). The primary immunization of female New Zealand White rabbits consisted of 100 nmol of conjugated peptide emulsified with complete Freund's adjuvant (Difco Laboratories, Michigan USA), 1 ml total volume was administered over five subcutaneous (s.c.) sites. Subsequent immunizations consisted of 50 nmol of conjugated peptide emulsified in incomplete Freund's adjuvant and administered over five s.c. sites. Pre-immune serum was collected from each rabbit before the immunization regime commenced, a test bleed was taken from each animal after it had received the primary immunization and two booster injections, which were given at fortnightly intervals.

Each serum was tested by ELISA for reactivity to the synthetic peptide against which it was raised and was subsequently tested against recombinant Smd1 and Smd2 for sensitivity, specificity and cross-reactivity in ELISA and Western blotting. Three sera are used in this study. MJL3, a serum raised against amino acid numbers 1–10 in the mature Smd1 peptide (AAKPMGITCD), MJL8, a serum raised against amino acid sequence numbers 18–27 in the mature Smd2 peptide (AHCLLLGKSG) and MJL10, a control serum from a rabbit immunized with all of the components used for immunizations but without any synthetic peptide.

Haemolymph collection and antigen preparation

Haemolymph was collected from twenty-five adult *S. calcitrans* by puncturing the interthoracic membrane with a 0.33 mm diameter needle attached to a 1 ml syringe (Micro-Fine +, Becton Dickinson, UK) and withdrawing haemolymph from the body cavity taking care to avoid contamination of haemolymph with gut tissue or contents. Saturated phenylthiocarbamide (2 µl) was added to the pooled haemolymph sample to prevent coagulation. Samples were centrifuged at 12 000 *g* for 3 min and the supernatant used for ELISA.

Midgut dissection and antigen preparation

Anterior midguts (proventriculus, thoracic midgut and reservoir), reservoir zones alone, opaque zones alone or lipoid zones alone were dissected into 154 mM NaCl (pH 7.2), frozen immediately in liquid nitrogen and stored at −80 °C until required. Groups of twenty-five samples were homogenized in 100 µl of 154 mM NaCl (pH 7.2) and centrifuged at 9000 *g* for 15 min. The supernatant was removed and used in subsequent experiments. Contents of the reservoir zone or lipoid zone lumen were collected by removing the midgut region as described above and gently applying pressure along the length of the tissue with a Micro-Fine + needle that had been bent at right angles to the syringe barrel. Care was taken not to damage the gut tissue, the absence of accidental damage was assessed microscopically. The expelled contents were collected and pooled from twenty-five flies then stored at −80 °C until required. Thawed gut lumen samples were used directly in experiments.

Enzyme-linked immunosorbent assay (ELISA)

Microtitre plates (Type M29A, F-form, PS microplates, Dynatech, West Sussex, UK) were coated with 5 µg/ml of the appropriate antigen in carbonate buffer (15 mM sodium carbonate, 35 mM sodium hydrogen carbonate, pH 9.6), overnight at 4 °C. Coating conditions were determined by checkerboard titration. One hundred microlitre volumes per well of antigen solution, primary and secondary antibodies and substrate were used throughout. The plates were washed three times in PBS + Tween-20 (PBS-Tween; 145 mM sodium chloride, 2 mM sodium dihydrogen orthophosphate, 4 mM disodium hydrogen phosphate, 0.05% Tween-20, pH 7.2) between each step. The plates were blocked with 200 µl/well of 5% skimmed milk powder (Marvel) in PBS-Tween for 2 h at ambient temperature. After washing, the plates were incubated with the appropriate rabbit antisera (diluted 1 : 5000 in PBS-Tween) for 2 h at room temperature. After further washing the plates were incubated with horseradish peroxidase-labelled goat anti-rabbit IgG conjugate (Nordic Immunology, Tilburg, the Netherlands; diluted 1 : 1000 in PBS-Tween) for 2 h at room temperature. Following further washing the extent of binding was measured colourimetrically after the addition of 300 µl of 2,2′-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium salt (300 µl of a 20-mg/ml solution diluted in 9.7 ml of 2.3% citric acid solution, pH 4.0) in the presence of hydrogen peroxide (10 µl of 30% v/v H_2O_2). Optical density was measured in an ELISA plate reader (Titertek Twinreader, version 2.01) at wavelength 405 nm (OD_{405nm}). The optical density of each antigen preparation was measured in triplicate for each antibody used.

In order to account for day-to-day and plate-to-plate variability and the different volumes of tissues and/or their contents analysed in this study, two calculations were used. The first was a correction factor that allowed direct comparison between plates, the second was the expression of corrected OD_{405nm} values as units per tissue per fly (UTF).

Correction factor calculation

Each plate contained positive reference peptides (against which the antisera were raised). If the mean optical density of positive reference wells was not exactly 1.0, the correction factor was applied. This correction factor was 1.0/mean optical density of reference positive wells. The mean OD_{405nm} for each antigen preparation was then multiplied by the correction factor obtained for that antibody.

Expression of units per tissue per fly (UTF)

The equation for this was: xy/5, where x = volume of sample collected and y = total protein expressed as µg/µl.

The corrected optical density was multiplied by the resulting factor, which was then divided by the number of flies used in the group (typically twenty-five) thus giving UTF. UTFs were also calculated for the control sera and subtracted from the UTFs obtained for anti-Smd1 and anti-Smd2 sera.

Tris-tricine SDS-PAGE and Western blotting

Proteins diluted in sample buffer (2 × sample buffer: 0.1 M Tris, 4% SDS, 5% 2-mercaptoethanol, 0.01% Coomassie blue G250, pH 6.8) and separated in Tris-tricine SDS-PAGE (Schagger & von Jagow, 1987) using a 16.5% T/3% C separating gel, a 10% T/3% C spacer gel and a 4% T/3% C stacking gel. Polypeptide molecular weight markers (Bio-Rad, UK) were simultaneously electrophoresed. After electrophoresis protein samples were stained with Coomassie Blue, silver stain or electrotransferred on to nitrocellulose paper (NCP) for Western blotting. Proteins separated by SDS-PAGE were electrotransferred to NCP (Hybond-C extra, Amersham Life Sciences, UK) according to previously published methods (Towbin *et al.*, 1979). After transfer, the marker

lanes were removed and stained with a colloidal gold solution (Protogold, British BioCell International) according to the manufacturers instructions. The remainder of the blots were blocked in 5% skimmed milk powder (Marvel) in Tween transblotting solution (TTBS, 20 mM Tris, 0.9% NaCl, 0.1% Tween-20, pH 7.2) for 2 h at ambient temperature. The blots were washed (three times for 5 min) in TTBS, cut into strips and probed with the appropriate rabbit antisera (diluted 1 : 75 in TTBS) for 2 h at room temperature. After subsequent washing (three times for 5 min in TTBS) the blots were incubated in horseradish peroxidase-labelled goat anti-rabbit IgG (diluted 1 : 1000 in TTBS, Nordic Immunology, Tilburg, the Netherlands) for 2 h at room temperature. After further washing the blots were developed by the addition of 4,chloro-1-naphthol (20 mg in 4 ml of methanol) in 16 ml of trans-blotting solution (TBS, 20 mM Tris, 0.9% NaCl, pH 7.2) containing hydrogen peroxide (10 µl of 30% v/v H_2O_2). The reaction was stopped by immersing the blots in distilled water.

Purification of > 26 kDa band and N-terminal sequencing

The > 26 kDa band was purified by preparative tris-tricine SDS-PAGE followed by excision of the correct band and electro elution of the protein of interest. During each PAGE run the ends of the protein band were cut from the gel and Western blots were carried out on each end of the gel with Smd1 antiserum. This enabled us to monitor that we had the correct band and allowed us to align the blotted ends with the gel in order to excise the band of interest. When we had a single band on tris-tricine SDS-PAGE we confirmed its identity by blotting and excised and electro-eluted the protein again. The purified protein was then subjected to tris-tricine SDS-PAGE and electro-transferred on to polyvinylidene fluoride membrane (PVDF membrane, Sigma) in glycine-free CAPS buffer, pH 11.0. The membrane was stained with Coomassie blue and sent for commercial sequencing (Alta Biosciences, University of Birmingham, UK). One amino acid was observed per sequence cycle.

cDNA library, cloning and sequencing

A *S. calcitrans* adult midgut specific cDNA library, estimated to contain 1.4×10^6 individual clones, was constructed in Lambda ZAP (Stratagene) according to the manufacturers instructions. Eight hundred midguts were used to make the library. The library was plated using *E. coli* XL-1 Blue. A degenerate sense primer was constructed (5′-GGACAATCYCCXCAYCA-3′) based on the Edman sequence information for the > 26 kDa protein. This degenerate primer and a universal M13–20 primer were then used in PCR with the midgut specific cDNA library to generate a ^{32}P-labelled probe for screening the library. pBluescript phagemids were excised *in vivo* from the lambda vector using ExAssist helper phage and plated using *E. coli* XLOLR (Stratagene). DNA sequencing was carried out using a Beckman CEQ 2000XL capillary sequencer.

RT-PCR

The primers 5′-CATTGCTACTGGACCAGA-3′ and 5′-GGACAATTTCCTCACCA-3′ were designed from regions of the > 26 kDa protease gene that are not highly conserved in homologous sequences selected by Blastx. Poly A$^+$ RNA was extracted from fifteen anterior midguts (cardiac, thoracic and reservoir zones), fifteen posterior midguts (opaque and lipoid zones), six fat bodies and the remains of two carcasses (approximately equal wet weights of tissue) from adult *S. calcitrans* using the Dynabeads system. One tenth of the extract was used in RT-PCR using the Access RT-PCR system (Promega) with one cycle of 45 min at 48 °C, one cycle of 2 min at 94 °C, forty cycles of 30 s at 94 °C, 1 min at 60 °C and 2 min at 68 °C. Finally we performed one cycle for 7 min at 68 °C. Controls omitting reverse transcriptase were used to check for genomic DNA contamination.

Zone inhibition assays

Antibacterial activity was estimated using zone inhibition assays utilizing either *Micrococcus luteus* or *E. coli* D31 (Lehane *et al.*, 1997). Diameters of zones were recorded following 24 h incubation at either 28 °C for *M. luteus* or 37 °C for *E. coli* D31. Midgut homogenates were filtered through a 0.2 µm nylon filter (Whatman) by centrifuging at 25 000 *g* for 15 min to remove any contaminating bacteria that possess antimicrobial activity (e.g. *Serratia marcescens*; J. V. Hamilton, unpublished results). Samples containing 4 midgut equivalents were incubated with either 0.9% saline or antibody (diluted in saline) for 15 min at ambient temperature before loading into the wells of the agarose plate. The plates were incubated at 30 °C for 16–24 h and the zones of inhibition measured.

Acknowledgements

This work was supported by grants from the Wellcome Trust and BBSRC.

References

Adrian, T.E. (1997) Production of antisera using peptide conjugates. *Neuropeptide Protocols. Meth Mol Biol* **73**: 239–249.

Altschul, S.F., Madden, T.1., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Blakemore, D., Lehane, M.J. and Williams, S. (1993) Cyclic-AMP can promote the secretion of digestive enzymes in *Stomoxys calcitrans*. *Insect Biochem Mol Biol* **23**: 331–335.

Brey, P.T., Lee, W.J., Yamakawa, M., Koizumi, Y., Perrot, S., Francois, M. and Ashida, M. (1993) Role of the integument in insect immunity: epicuticular abrasion and induction of cecropin synthesis in cuticular epithelial cells. *Proc Natl Acad Sci USA* **90**: 6275–6279.

Dimopoulos, G., Richman, A., Muller, H.M. and Kafatos, F.C. (1997) Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc Natl Acad Sci USA* **94**: 11508–11513.

Gabay, J.E. and Almeida, R.P. (1993) Antibiotic peptides and serine protease homologs in human polymorphonuclear leukocytes: defensins and azurocidin. *Curr Opin Immunol* **5**: 97–102.

Hao, Z., Kasumba, I., Lehane, M.J., Gibson, W.C., Kwon, J. and Aksoy, S. (2001) Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc Natl Acad Sci USA* **98**: 12648–12653.

Hoffmann, J.A. and Reichhart, J.M. (1997) *Drosophila* immunity. *Trends Cell Biol* **7**: 309–316.

Jordao, B.P., Lehane, M.J., Terra, W.R., Ribeiro, A.F. and Ferreira, C. (1996) An immunocytochemical investigation of trypsin secretion in the midgut of the stable fly, *Stomoxys calcitrans*. *Insect Biochem Mol Biol* **26**: 445–453.

Kato, K., Kishi, T., Kamachi, T., Akisada, M., Oka, T., Midorikawa, R., Takio, K., Dohmae, N., Bird, P.I., Sun, J.R., Scott, F., Miyake, Y., Yamamoto, K., Machida, A., Tanaka, T., Matsumoto, K., Shibata, M. and Shiosaka, S. (2001) Serine proteinase inhibitor 3 and murinoglobulin I are potent inhibitors of neuropsin in adult mouse brain. *J Biol Chem* **276**: 14562–14571.

Kelly, J.M., O'Connor, M.D., Hulett, M.D., Thia, K.Y.T. and Smyth, M.J. (1996) Cloning and expression of the recombinant mouse natural killer cell granzyme Met-ase-1. *Immunogenetics* **44**: 340–350.

Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis. *Ann Rev Biochem* **46**: 331–358.

Lehane, M.J. (1991) *Biology of Blood-Sucking Insects*, 1st edn. Chapman & Hall London, UK.

Lehane, M.J. (1997) Peritrophic matrix structure and function. *Ann Rev Entomol* **42**: 525–550.

Lehane, S.M., Assinder, S.J. and Lehane, M.J. (1998) Cloning, sequencing, temporal expression and tissue-specificity of two serine proteases from the midgut of the blood-feeding fly *Stomoxys calcitrans*. *Eur J Biochem* **254**: 290–296.

Lehane, M.J., Wu, D. and Lehane, S.M. (1997) Midgut-specific immune molecules are produced by the blood-sucking insect *Stomoxys calcitrans*. *Proc Natl Acad Sci USA* **94**: 11502–11507.

Maudlin, I. (1991) Transmission of African trypanosomiasis: interactions among tsetse immune system, symbionts and parasites. *Adv Dis Vector Res* **7**: 117–149.

Medzhitov, R. and Janeway, C.A. (1997) Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**: 295–298.

Muller, H.M., Catteruccia, F., Vizioli, J., DellaTorre, A. and Crisanti, A. (1995) Constitutive and blood meal-induced trypsin genes in *Anopheles gambiae*. *Exp Parasitol* **81**: 371–385.

Munks, R.J.L., Hamilton, J.V., Lehane, S.M. and Lehane, M.J. (2001) Regulation of midgut defensin production in the blood-sucking insect *Stomoxys calcitrans*. *Insect Mol Biol* **10**: 561–571.

Nakajima, Y., Tsuji, Y., Homma, K. and Natori, S. (1997) A novel protease in the pupal yellow body of *Sarcophaga peregrina* (flesh fly). Its purification and cDNA cloning. *J Biol Chem* **272**: 23805–23810.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**: 1–6.

Noriega, F.G. and Wells, M.A. (1999) A molecular view of trypsin synthesis in the midgut of *Aedes aegypti*. *J Insect Physiol* **45**: 613–620.

Sayers, T.J., Brooks, A.D., Ward, J.M., Hoshino, T., Bere, W.E., Wiegand, G.W., Kelley, J.M. and Smyth, M.J. (2001) The restricted expression of granzyme M in human lymphocytes. *J Immunol* **166**: 765–771.

Schagger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**: 368–379.

Scott, F.L., Hirst, C.E., Sun, J.R., Bird, C.H., Bottomley, S.P. and Bird, P.I. (1999) The intracellular serpin proteinase inhibitor 6 is expressed in monocytes and granulocytes and is a potent inhibitor of the azurophilic granule protease, cathepsin G. *Blood* **93**: 2089–2097.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* **76**: 4350–4354.

Tsuji, Y., Nakajima, Y., Homma, K. and Natori, S. (1998) Antibacterial activity of a novel 26-kDa serine protease in the yellow body of *Sarcophaga peregrina* (flesh fly) pupae. *FEBS Lett* **425**: 131–133.

Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J.M., Lemaitre, B., Hoffmann, J.A. and Imler, J.L. (2000) Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia. *Immunity* **13**: 737–748.

Vernick, K.D., Fujioka, H., Seeley, D.C., Tandler, B., Aikawa, M. and Miller, L.H. (1995) Plasmodium gallinaceum: *a refractory mechanism of ookinete killing in the mosquito*, Anopheles gambiae. *Exp Parasitol* **80**: 583–595.

Wilson, C.L., Ouellette, A.J., Satchell, D.P., Ayabe, T., Lopez-Boado, Y.S., Stratman, J.L., Hultgren, S.J., Matrisian, L.M. and Parks, W.C. (1999) Regulation of intestinal alpha-defensin activation by the metalloproteinase matrilysin in innate host defense. *Science* **286**: 113–117.

Vol. 175, No. 3, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 29, 1991 Pages 784-794

EXPRESSION OF AN AUTOPROCESSING CAT-HIV-1 PROTEINASE FUSION PROTEIN: PURIFICATION TO HOMOGENEITY OF THE RELEASED 99 RESIDUE PROTEINASE

Douglas S.Montgomery[1], Onkar M.P.Singh[1], Norman M.Gray[2], Colin W.Dykes[1], Malcolm P. Weir[1] and Adrian N.Hobden[1]

[1]Department of Genetics and [2]Virology
Glaxo Group Research Ltd, Greenford, Middlesex, United Kingdom

Received January 7, 1991

The 99 residue human immunodeficiency virus type 1 proteinase has been expressed in *Escherichia coli* as part of an autocleaving fusion protein. Expression of the fusion protein is toxic to the host cells, however yields of the released proteinase have been improved by optimising induction and harvest times to increase culture biomass, and decrease degradation of the proteinase. Soluble proteinase was extracted from these cells by a simple and highly efficient three step process. N-terminal sequence analysis confirms that the enzyme preparation is highly pure and correctly autoprocessed. The proteinase cleaves peptide substrate lGCTLNFPISPIETV between F and P at pH 6.0 with a Km of 310μM and a Kcat of 14s^{-1}. The enzyme is sensitive to its ionic environment, showing stimulation of activity at high salt concentrations, and shows a pH optimising 5.5.

The role of human immunodeficiency virus type 1 (HIV-1) in the onset of acquired immunodeficiency syndrome (AIDS) has been clearly demonstrated [1,2]. The HIV-1 proteins encoded by GAG, POL and ENV genes are expressed as precursor molecules which are then processed by proteolytic cleavage to yield the structural proteins and enzymes required for viral replication [3]. Processing of the GAG and GAG-POL precursor proteins is mediated by a virally encoded proteinase [4]. This 99 amino acid residue proteinase, encoded within the 5' end of the POL open reading frame [5], is released from the GAG-POL precursor by an autocatalytic processing event which cleaves at two flanking Phe-Pro sites [6-11]. The HIV-1 proteinase, like those encoded by other retroviruses, shows homology with the N- and C-terminal domains of the family of aspartyl proteinases [12-14], particularly within the active site region Asp-Thr-Gly motif. The sensitivity of the enzyme to pepstatin [7, 15-17] confirms that this is a true member of the aspartyl proteinase family. The 11kDa POL proteinase contains only one half of the residues which contribute to the classical aspartyl proteinase active site, however the recently determined crystal structures of the proteinases from Rous sarcoma virus and HIV-1 show that they form dimeric

ABBREVIATIONS: IPTG, isopropyl-β-D-thiogalactopyranoside; EPNP, 1,2-epoxy-(4-nitrophenoxy) propane; DAN, N-diazoacetylnorleucine; MES, 4-morpholine ethane sulphonic acid.

0006-291X/91 $1.50

structures [18-20]. Processing of the GAG and POL precursors is specifically prevented by mutation of the proteinase active site aspartate (POL residue Asp93) to alanine, asparagine or threonine [6,10,16,21]. The latter mutation made in a proviral construct resulted in the production of immature non-infectious particles, thus showing that the HIV-1 proteinase is essential for viral reproduction and therefore a valid target for viral chemotherapy.

In our initial attempts to express the HIV-1 proteinase in *E.coli* we were unable to detect the proteinase by Western blot. This was almost certainly due to the rapid arrest of culture growth, upon induction of proteinase expression, preventing accumulation of the proteinase. In order to increase the expression we have fused the proteinase to the highly expressed chloramphenicol acetyl transferase (CAT) gene in the plasmid pTCX2TT [24] and altered the induction and harvest times to increase production of the proteinase to levels that allow purification of the enzyme to near homogeneity.

MATERIAL AND METHODS

Plasmid Construction

M13mp11 was modified by insertion of a synthetic linker (Figure 1a) into the multiple cloning site. The BglII-EcoRI POL fragment of lambda phage clone BH10[5] was inserted between the BamHI and EcoRI sites of the modified M13 vector. Translation termination codons and a SalI site were introduced at the 3' end of the predicted POL proteinase coding sequence by SDM. These manipulations created on NcoI-SalI flanked sequence (designated MatP) encoding residues 1-167 of the POL orf, preceded by a Met-Gly dipeptide. The NcoI-SalI MatP fragment was fused with the 5' end of the CAT gene in plasmid pTCX2TT [24] via a synthetic EcoRI-NcoI linker (5'-AATTCGGATCCAACGG-3' and 5'-CATGCCGTTGGATCCG-3'). The resultant plasmid pTCEMatP1 (Figure 1c) encodes a 27KDa CAT-MatP fusion protein under control of the IPTG inducible *Tac* promoter. Site directed mutagenesis was used to change the proteinase active site residue Asp93 (GAT) to Asn (AAT). The mutated form of MatP (designated MatP(D93N)) was fused to the CAT gene as described above to generate plasmid pTCEMatP1(D93N). Cultures of *E.coli* RB791 carrying the expression plasmids were grown in minimal media as described [24].

Proteinase Extraction, Purification and Analysis

Cells were harvested from 10 one litre, shake flask, cultures by centrifugation at 10,000g for 20 minutes at 4°C. The proteinase was extracted from the cells, essentially as described [7]. Cells were resuspended at 2g (wet weight)/ml in buffer A (50mM MES, 2mM EDTA, 0.01% Triton X-100, pH 6.0) then extensively sonicated. The total cell lysate was diluted with 10 vol/vol cold acetone, left at -70°C for 30 minutes then centrifuged at 10,000g for 30 minutes. The pellet was dried under nitrogen and stored at -70°C until required. It was solubilised with buffer A containing 0.5M NaCl and 40% glycerol and centrifuged at 100,000g for 30 minutes. The supernatant was diluted 10 fold into buffer B (50mM MES, 2mM EDTA, pH 6.0) and immediately loaded, at 4ml/min, onto a Mono S (HR10/10) column. The enzymic activity was recovered by a 0-1M NaCl gradient (0-0.2M in 48ml, isocratic at 0.2M to 96ml, 0.2-0.6M to 224ml and 0.6-1.0M to 272ml) in buffer B. The active fractions were pooled and diluted with equal volume of 3.2M ammonium sulphate in buffer B. This was applied to the alkyl-superose (HR 5/5) column at 1.0ml/min and eluted, at 0.5ml/min, by applying a 1.6-0M ammonium sulphate gradient (1.6-0M in 4ml and OM isocratic to 6.0ml) in buffer B. 1ml fractions were collected into tubes containing glycerol (50% v/v) and kept in ice. The pooled material was then applied to a gel permeation column (2.6cm x 70cm) containing superdex G75 matrix. The column was eluted with buffer B containing 0.5M NaCl and 10% glycerol. Fractions containing proteinase activity were pooled, the glycerol concentration raised to 50%, and stored at -70°C. Proteinase activity was determined by incubating 1-5 μl of protein (0.05-2 μg) with 150μM substrate peptide (IGCTLNFPISPIETV) in buffer A

containing 2M NaCl at 37° C for up to 5 minutes in a final volume of 200µl. The reaction was terminated by adding 4µl of 80% acetic acid and boiling the mixture for 5 minutes. After centrifugation at 10,000g for 5 minutes, 100µl of the sample was applied to a reverse phase chromatography (PepRpc) column running at 1.0ml/min. Both substrate and products were resolved by an elution gradient of 0.1% TFA/water and 0.1% TFA/acetonitrile (0-15% acetonitrile in 1.5ml, isocratic at 15% to 3.5ml, 15-40% to 13.5ml and isocratic at 40% 14.5ml) and quantified by integration of the peak areas at 214nm. The kinetic data was analysed using the ENZFITTER software package.

N-terminal sequence analysis was by automated Edman degradation using a gas phase sequencer (Applied Biosystems Model 477A). Identification and quantification of the phenylthiohydantoin derivatives was performed on line using an Applied Biosystems Model 120A PTH amino analyser with external quantification. The protein sample (200pmole in 50% acetonitrile, 0.1%TFA) was loaded on a pre-cycled polybrene filter and used for 15 cycles of sequencing. Protein concentration in the crude extract and the partially purified enzyme preparation was determined using the Bicinchoninic acid reagent [25]. Purified enzyme protein concentration was determined from its absorbance at 280nm using a value of E 1% = 11.4 which was calculated from the extinction coefficients of both Trp and Tyr residues.

RESULTS AND DISCUSSION

The CAT-HIV-1 proteinase gene fusion vector pTCEmatP1 (Fig 1c) encodes a 27kDa fusion protein containing the N-terminal 72 residues of CAT and the first 167 residues of the HIV-1 POL open reading frame, including the 99 residue proteinase, and the autocleavage site (Phe 68-Phe 69) at its N-terminus. The fusion protein was predicted to be capable of autocleavage to release the mature 11kDa proteinase. Induction of fusion protein expression resulted in arrest of culture growth (Fig 2). Cultures induced 1 to 4 hours after inoculation ceased to grow within 2-3 hours. Induction after 5 hours coinsided with the onset of stationary phase growth of the uninduced control culture. The specific proteinase activity in the soluble extract of cultures induced after 1 to 4 hours was similar, but approximately two fold higher than that observed in soluble extracts of the culture induced after 5 hours. No growth retardation was observed in cultures induced to express the mutated fusion protein from pTCEmatP1 (D93N), and only background proteinase activity was detected in the soluble fractions of these extracts.

SDS-PAGE analysis was carried out on whole cell extracts from cultures of pTCEmatP1 and pTCEmatP1(D93N) induced after 4 hours (A_{550} of 1.4-1.6) and harvested 30 minutes after induction. No specifically induced bands were visible on Coomassie blue stained gels of the pTCEmatP1 extracts (Fig 3a). However, extracts of pTCEMatP1(D93N) cultures revealed a specifically induced protein of 27kDa, which corresponds to the predicted size of the fusion protein. Western blot analysis with the proteinase antisera (Fig 3b) showed that both plasmids inducibly express an immunoreactive 27kDa protein, although it is present in greater quantity in the pTCEmatP1(D93N) extract, as would be expected from the stained gel result. A second immunoreactive protein, which appears exclusively in the induced pTCEmatP1(D93N) extract with a molecular weight of approximately 24kDa, is thought to be a degradation product from the 27kDa fusion protein. An induced immunoreactive protein of 11kDa was found in the pTCEmatP1 extracts

a

```
     EcoRI* NcoI                BamHI        EcoRI                SalI
5'  AATTGCCATGGGATTTTTTAGGGAGGATCCTCTAGATGAATTCCCTAATAACTAACTAAG      3'
      CGGTACCCTAAAAAATCCCTCCTAGGAGATCTACTTAAGGGATTATTGATTGATTCAGCT

          MetGlyPhePheArgGluAsp          IleProEndEnd End End
```

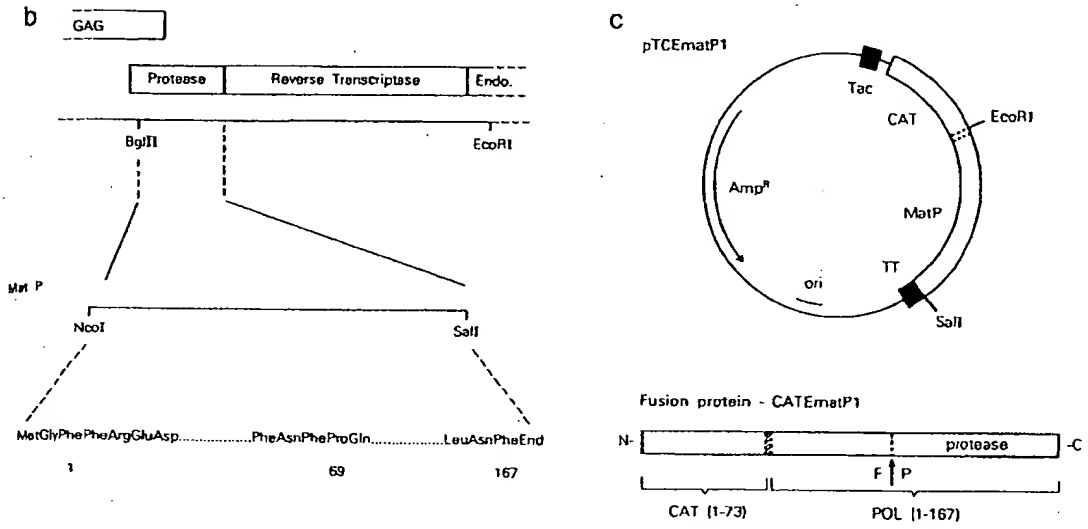


Figure 1. a) Synthetic linker used to modify M13 mp11 prior to insertion of Bg1II-EcoRI POL fragment.

b) The relationship of the NcoI-SalI matP gene fragment to the HIV-1 GAG and POL open reading frames. The relevant region of the amino acid sequence encoded by MatP are indicated (numbering corresponds to the POL orf).

c) Map of plasmid pTCEmatP1 and a schematic of the fusion protein it encodes. The Phe-Pro autoprocessing site at the N terminus of the 99 residue protease is indicated in the fusion protein.

but not in the pTCEmatP1(D93N) extracts and was therefore believed to be the released 99 residue HIV-1 proteinase. These results indicate that the fusion protein expressed from pTCEmatP1 is processed to release the 11kDa proteinase, probably by cleavage at the Phe68-Pro69 junction, whereas the 27kDa fusion protein containing the mutated proteinase from pTCEmatP1(D93N) does not release the 11kDa protein, but accumulates in the cell. This latter result suggests that the specific processing event is carried out by the HIV-1 proteinase rather than an *E.coli* proteinase. We would normally expect the two gene fusion to be expressed at the same rate since they differ by a single base pair, however, while the mutated fusion protein is visible on Coomassie blue stained gel, the detectable levels of unmutated fusion protein and released proteinase are considerably lower. This probably reflects the decreased growth rate of cultures induced to express the proteinase fusion protein (Fig 2) noted above. The toxic effect and low yield of HIV-1 proteinase in *E.coli* has been reported previously [16,26] and utilised as a screen for inactivating mutations in the proteinase gene [26]. The poor accumulation of released proteinase is probably the result of both reduced expression of the fusion protein and increased degradation of the released proteinase and/or fusion protein in

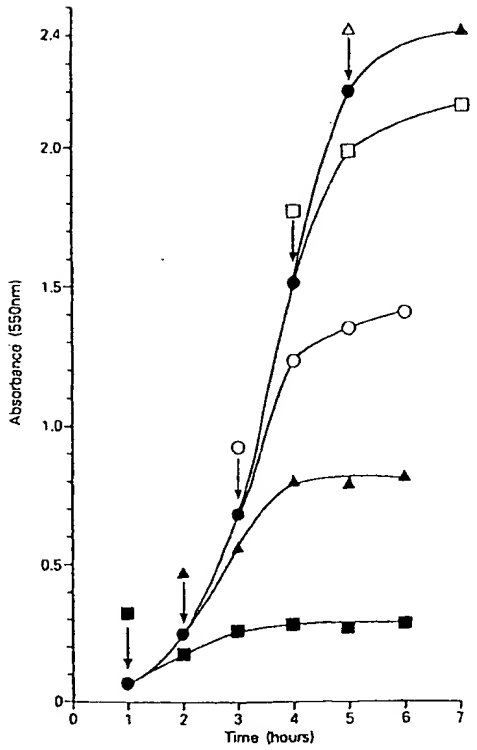


Figure 2. To assess the effect of expression from pTCEmatP1 in *E.coli* RB791 a series of cultures were induced with IPTG (1mM) at different times after inoculation as indicated by arrows (●No IPTG). Culture growth was monitored at 550nm and cultures were harvested 2 hours after induction.

these stressed cells. By sampling induced cultures at intervals post-induction we found that accumulation of the 11kDa proteinase peaks 15-45 minutes after addition of IPTG (data not shown). This corresponds to a period when the culture is still growing relatively rapidly (Fig 2).

To maximise accumulation of the released proteinase, larger scale cultures, for purification of the 11kDa enzyme were therefore induced at an absorbance of 1.4-1.6 and harvested within 45 minutes. 10g of cells harvested from shake flask cultures were used for the purification of the released proteinase. Acetone precipitation, detergent solubilisation and the subsequent centrifugation resulted in the removal of a large amount of cellular debris as well as the lipid. The extract was diluted 10 fold before applying it to the cation exchange column. Elution with a NaCl gradient resulted in over 80% of the activity being displaced at around 0.3M NaCl (Fig 4a). Fractions containing proteinase activity were pooled and then diluted (1:1) with buffer A containing 3.4M ammonium sulphate and loaded on the alkyl-Superose matrix. A considerable amount of UV absorbing material did not bind to this column and after washing with 10 column volumes, the column was developed with a decreasing linear gradient of ammonium sulphate (1.7-0.0M). 100% of the activity loaded was recovered in fractions 4 and 5 (Fig 4b). These were pooled and further chromatographed on the Superdex G75 gel permeation column. The enzyme appeared to elute with

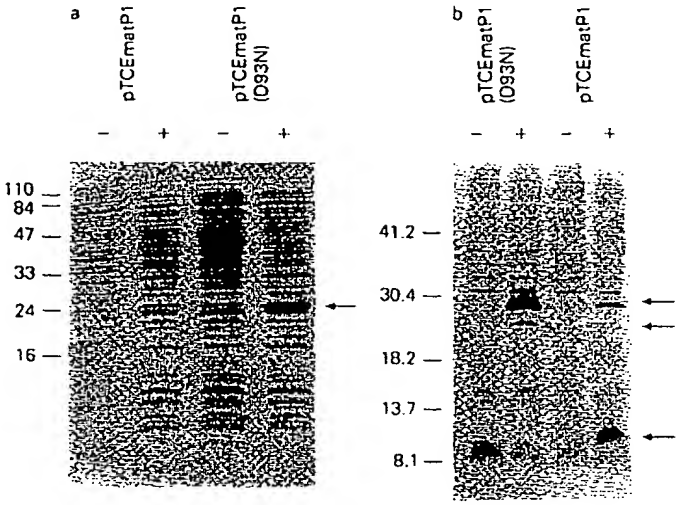


Figure 3. Gel analysis of whole cell extracts from pTCEmatP1 and pTCEmatP1 (D93N) cultures + and - induction with IPTG. Extracts were prepared from 1-2ml culture. Cell pellets were resuspended in 1 x gel loading buffer at 10 A_{550} units per ml, boiled for 5 mins and 10μl loaded for analysis on a) Coomassie blue stained 20% SDS-PAGE [33] and b) Western blot immuno-stained [34] with proteinase specific peptide antisera. Rabbit antisera was raised to peptides corresponding to HIV-1 POL residues 93-99 and 107-113. Arrows indicate induced bands. Sizes of molecular weight markers are indicated in kDa.

a retention volume consistent with it existing as a monomer of near 10kDa in size (Fig 4c), with >80% of the enzymic activity being recovered. The total amount of protein recovered after this final step was 200μg and the specific activity was determined to be 45μmol/min/mg of protein. Increase in NaCl concentration to 2M resulted in the enzymic activity eluting from the Superdex G75 column at a volume consistent with it being dimeric. However, the separation was carried out at 0.5M NaCl to remove larger contaminating proteins. This effect of increasing ionic strength on the enzyme dimerisation and catalysis is discussed elsewhere [27].

Fraction containing proteinase activity were pooled and re-chromatographed on the alkyl-superose matrix to concentrate the enzyme. SDS-PAGE showed that the preparation was over 90% pure (Fig 5). The amino acid analysis (Table 1) was consistent with the predicted composition of the proteinase. 15 cycles of N-terminal analysis by gas phase sequencing showed the following sequence PQITLWQRPLVTIK, consistent with the known proteinase sequence and indicating that the proteinase is released by cleavage at the predicted F-P cleavage site. The purified enzyme was stored at -70° C in buffer at pH 6.0. Autodegredation upon storage has been reported [28] but in this case the enzyme was stored in acetic acid (pH 4.0). The pH/activity profile displayed by our proteinase (Fig 6) is similar to that reported [18,29]. Maximum activity was detected at pH 5.5. The Km for the peptide, IGCTLNFPISPIETV, increases from 180μM at pH 5.0 to >1mM at pH 7.0. The effect of varying the pH on the inhibition with acetyl-pepstatin and the renin inhibitor H-261 has

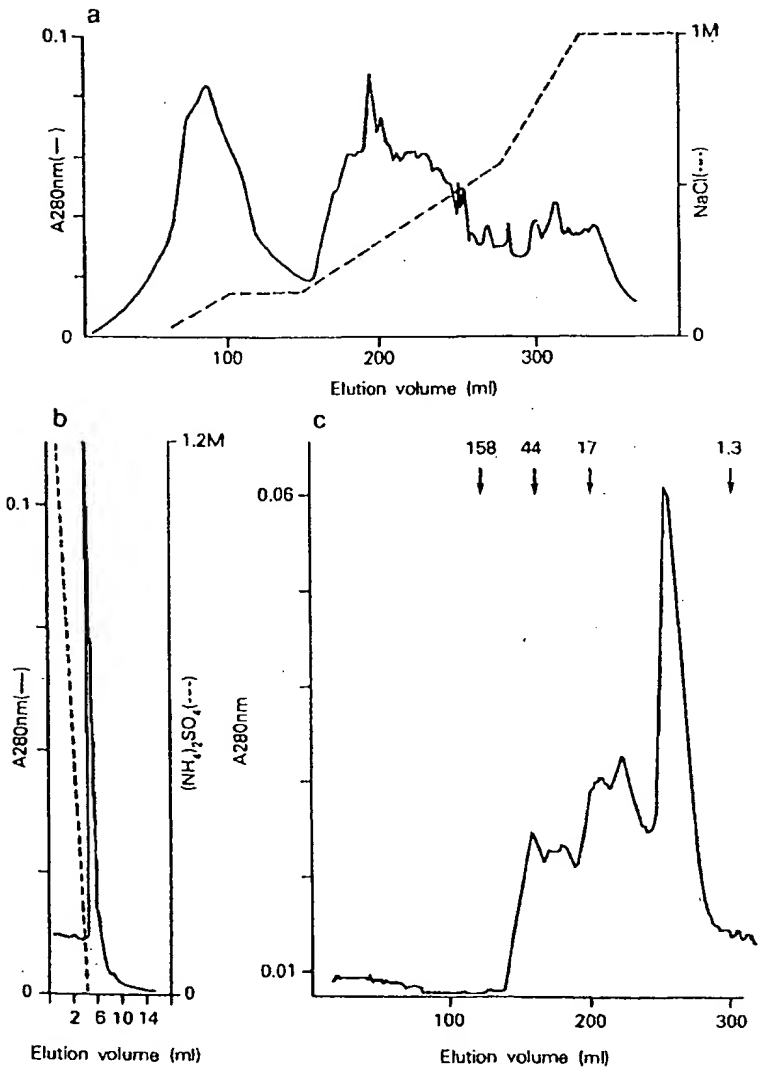


Figure 4. Column Chromatography used to purify the released HIV-1 proteinase. These columns were loaded and run as described in Materials and Methods.

a) FPLC mono S (HR 10/10) column profile. 20ml fractions were collected. Fractions 7 to 12 (elution volume 140 to 240ml) contained >80% of the detected proteinase activity.

b) Alkyl-superose (HR 5/5) column profile. The proteinase activity recovered from this column corresponded to the major protein peak (elution volume 5 to 8ml).

c) Superdex G75 column profile. This column was calibrated with BioRad gel filtration standards. The elution volume of these standards are indicated (kDa). All the proteinase activity eluted with the major peak between 1.3 and 17kDa.

also been reported [17]. Interestingly, increasing the pH from 4.7 to 7.0 produced a decreased binding of both compounds, with Ki for the first increasing 50 fold and the second only 5 fold. This dramatic effect may be the result of the ionisation property of the His residue in H-261 and the C-

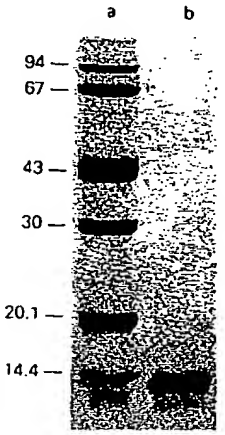


Figure 5. Silver stained 20% SDSPAGE gel of purified HIV-1 proteinase. Lane a. molecular weight markers (kDa) Lane b. 300ng purified proteinase. The purified proteinase was TCA precipitated and washed with ethanol prior to electrophoresis [33] and silver staining [35].

terminus carboxyl group of acetyl-pepstatin. Hence the ionisation of the substrate or inhibitor as well as that of the catalytic Asp residues may have a major influence on their binding. The purified enzyme has been used to compare the cleavage of peptides containing the Y-P and M-M sites. The Km values, at pH 6.0, for peptides VSQNYPIVQNIG, ATMMQRG and IGCTLNFPISPIETV were 230, 504 and 312μM respectively. These results were obtained under conditions which displayed near maximal proteinase activity ie. the enzyme was assayed in the presence of 2M NaCl. The Km values, whilst reflecting slightly different peptide sequences, are at least an order of magnitude lower than those previously reported [30,31].

From the kinetic analysis of peptide IGCTLNFPISPIETV hydrolysis, a turnover number of near $14s^{-1}$ was calculated. Incubation of our purified enzyme with 10mM EPNP or DAN (EPNP and DAN are Asp modifying reagents) and 0.1mM Cu^{2+} resulted in both cases in 50% inhibition of the activity. Cu^{2+} alone caused 20% inhibition. This behaviour is consistent with HIV proteinase belonging to the aspartic proteinase family. 1mM iodoacetic acid and p-chloromercuribenzoate (thiol modifying reagents) each caused 75-80% inhibition. There are two thiol groups in the 99 residue proteinase, at position 66 and 95. Cys 95 appears buried at the dimerisation interface [20], therefore it is likely that modification of its thiol group results in the near complete abolition of proteinase activity by preventing dimerisation. This suggests that the subunits are in dynamic equilibrium under the experimental conditions, since Cys 95 would not be available for modification in the dimer.

CONCLUSION

We have overcome the toxic effect of expressing HIV-1 proteinase in *E.coli* by using the highly expressed CAT gene to rapidly produce the proteinase as an autoprocessing fusion protein and

Table 1 Amino acid composition analysis of the purified proteinase. The protein was concentrated, and the buffer constituents removed by reverse phase chromatography (ProRpc HR 5/10). The protein was eluted by applying a linear gradient (0-100%) of 0.1% TFA/water and 0.1% TFA/acetonitrile. The material eluting at around 52-54% acetonitrile was freeze dried and hydrolysed in the HCl vapour phase using a Waters Pico Tag Workstation. Phenylisothiocyanate (PITC) derivatives, with aminobutyric acid as internal standard, were separated by reverse phase chromatography essentially using the manufacturers instructions (Waters-Millipore, UK).

RESIDUE	NUMBER EXPECTED	NUMBER DETECTED
D/N	7	5.8
E/Q	10	9.1
S	1	2.0
G	13	13.1
H	1	0.7
R	4	4.4
T	8	7.6
A	3	4.1
P	6	6.2
Y	1	0.8
V	6	6.0
M	2	2.1
I	13	9.6
L	12	11.8
F	2	1.8
K	6	4.9

altering induction and harvest times to maximise accumulation of the released proteinase. The enzyme has purified to near homogeneity and is found to be highly active against various peptide substrates containing predicted HIV cleavage sites. The enzyme is inhibited by DAN, EPNP, IVA-pepstatin, acetyl-pepstatin and acetyl-pepstatinamide. Its sensitivity to iodoacetic acid and p-chloromercurial benzoate may well be due to the presence of a thiol group (Cys95) buried at the dimerisation interface which becomes modified when exposed in the free monomer. The modification of the thiol group in HIV proteinase may therefore result in destabilisation of dimerisation. Kinetic analysis of the cleavage of peptide IGCTLNFPISPIETV demonstrates that the purified enzyme is efficient with a Kcat/Km of around $10^5 M^{-1}s^{-1}$. Variations in the kinetic constants reported by others have been discussed [29]. It is likely that different assay conditions, pH

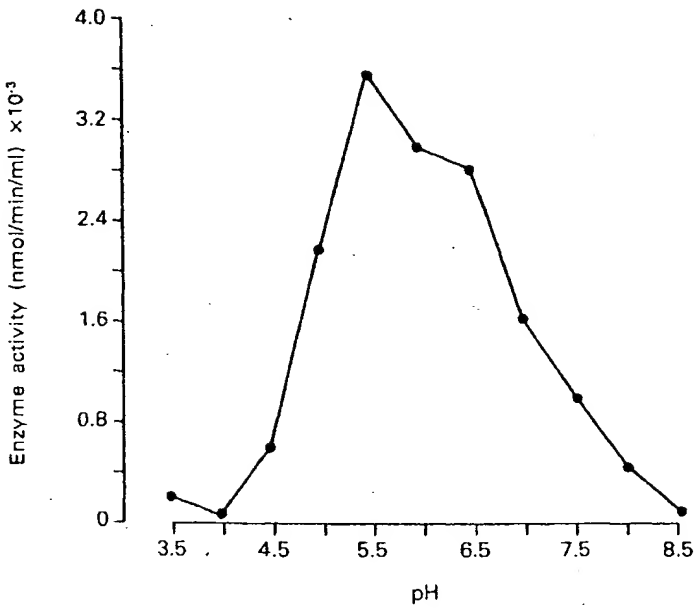


Figure 6. The effect of pH on the initial activity (V_0) of the purified HIV-1 proteinase. The enzyme was assayed as described in Materials and Methods with 2mM EDTA in the reaction mixture buffered with 50mM glycine-HCl (pH 3.0-3.5), 50mM Na-acetate (pH 4.0-4.5), 50mM Mes (pH 5.0-7.0) or 50mM Tris-HCl (pH 7.5-8.5). 150μM peptide IGCTLNFPISPIETV was used as substrate.

and ionic strength have a strong influence on this. Any analysis of the effect of ionic strength on the catalysis in terms of increased binding of substrate [32] must also take into account the influence of ionic strength or pH on dimerisation [27].

ACKNOWLEDGMENTS

We would like to thank Ms J Sparks, Ms E M J Roud-Mayne and Mr A C Marshall for technical assistance. Dr G Turcatti and Dr P Wingfield (Glaxo Institute of Molecular Biology, Geneva, Switzerland) for N-terminal sequencing. Mr B A Coomber, Dr C Christodoulou, Dr J Kitchen and Dr P Seal (GGR, Greenford) for in house synthesis of oligonucleotides and peptides.

REFERENCES

1. Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Charmeret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983) Science, 220, 868-870.
2. Gallo, R.C., Salahuddin, S.Z., Popvic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P. and Markham, P. (1984) Science, 224, 500-502.
3. Coffin, J.M. (1982) in Molecular Biology of Tumor Viruses, RNA Tumor viruses, Eds Weiss, R., Teich, N., Varmus, H., and Coffin, J. (Cold Spring Harbor Lab. Cold Spring Harbor, NY) pp 261-368.
4. Wong-Staal, F., and Gallo, R.C. (1985) Nature, 317, 395-403.
5. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R. Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C., and Wong-Staal, F. (1985) Nature 313, 277-284.

6. Mous, J., Heimer, E.P., & Le Grice, S.F.J. (1988) J. Virol. 62 1433-1436.
7. Hansen, J., Billich, S., Shulze, T., Sukrow, S and Moelling, K. (1988) EMBO J. 7, 1785-1791.
8. Debouck, C., Gorniak, J.G., Stickler, J.E., Meek, T.D., Metcalf, B.W. and Rosenberg, M. (1987) Proc. Natl. Acad. Sci. USA, 84, 8903-8906.
9. Lillehoj, E.P., Salazar, F.H.R., Mervis, R.J., Raum, M.G., Chan, H.W., Ahmad, N. and Vankatesan, S. (1988) J. Virol. 62 3053-3058.
10. Le Grice, S.F.J., Mills, J. and Mous, J. (1988) EMBO J. 7, 2547-2553.
11. Graves, M.C., Lim, J.J., Heimer, E.P. and Kramer, R.A. (1988) Proc. Natl. Acad. Sci. USA 85, 2249-2453.
12. Toh, H., Ono, M., Saigo, K. and Miyata, T. (1985) Nature 315, 691.
13. Pearl, L.H. and Taylor, W.R. (1987) Nature 328, 482
14. Pearl, L.H. and Taylor, W.R. (1987) Nature 329, 351-354.
15. Seelmeier, S., Schmidt, H., Turk, V. and Von Der Helm, K. (1988) Proc. Natl. Acad. Sci. USA 85, 6612-6616.
16. Darke, P.L., Leu. C-T., Davis, L.J., Heimbach, J.C., Diehl, R.E., Hill, W.S., Dixon, R.A.F. and Sigal, I.S. (1989) 264, 2307-2312.
17. Miller, M., Sathynarayanar, B.K., Wlodawer, A., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Schneider, J. and Kent, S.B.H. (1989) Science 246, 1149-1152.
18. Richards, A.D., Roberts, R., Dunn, B.M., Graves, M.C. and Kay, J. (1989) FEBS Letts. 247, 113-117.
19. Miller, M., Jaskoliski, M., Rao, M.J.K., Leis, J. and Wlodawer, A. (1989) Nature 337, 576-579.
20. Navia, M., Fitzgerald, P.M.D., McKeever, B.M., Leu, C-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L. and Springer, J.P. (1989) Nature 337, 615-620.
21. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimback, J.C., Dixon, R.A.F., Scolnick, E.M. and Sigal, I.S. (1988) Proc. Natl. Acad. Sci. USA 85, 4686-4690.
22. Schneider, J and Kent, S.B.H. (1988) Cell 54, 363-368.
23. Nutt, R.F., Brady, S.F., Darke, P.L., Ciccarone, T.M., Colton, C.D., Nutt, E.M., Rodkey, J.A., Bennett, C.D., Waxman, L.H., Sigal, I.S., Anderson, P.S. and Veber, D.F. (1988) Proc. Natl. Acad. Sci. USA 85, 7129-7133.
24. Dykes, C.W., Bookless, A.B., Coomber, B.A., Noble, S.A., Humber, D.C. and Hobden, A.N. (1988) Eur. J. Biochem. 174, 411-416.
25. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.C., Gartner, F.H., Provenzano, M.D., Fuyimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. (1985) Anal. Biochem. 150, 76-85.
26. Baum, E.Z., Bebernitz, G.A. and Gluzman, Y. (1990) Proc. Natl. Acad. Sci. USA 87, 5573-5577.
27. Singh, O.M.P., Roud-Mayne, E.M.J. and Weir, M.P. (1990) in Retroviral Proteases: Control of Maturation and Morphogenesis. Ed. LH Pearl, McMillan Press UK in press.
28. Stickler, J.E.,Gormack, J., Dayton, B., Meek, T., Moore, M., Magaard, V., Malinowski, N. and Debouk, C. (1989) Protein: Structure Function and Genetics 6, 139-154
29. Tomarseli, A.G., Olsen, M.K., Hui, J.O., Staples, D.J., Sawyer, T.K., Heinrikson, R.L. and Tomich, C-S.C. (1990) Biochemistry 29, 264-269.
30. Meek, T.D., Dayton, B.D., Metcalf, B.W., Dreyer, G.B., Strickler, J.E., Gorniack, J.G., Rosenberg, M., Moore, M.L., Magaard, V.W. and Debouk, C. (1989) Proc. Natl. Acad. Sci. USA 86, 1841-1845.
31. Moore, M.L., Bryan, W.M., Fakhouri, S.A., Magaard, V.W., Huffman, W.F., Dayton, B.D., Meek, T.D., Hyland, L., Dreyer, G.B., Metcalf, B.W., Strickler, N.E., Gorniak, J.G. and Debouk, C. (1989) Biochem. Biophys. Res. Communs. 156, 297-303.
32. Richards, A.D., Phylip, L.H., Farmarie, W.G., Scarborough, P.E., Alvarez, A., Dunn, B.M., Hirel, P.H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V. and Kay, J (1990) J. Biol. Chem. 265, 7733-7736.
33. Laemmli, U.K. (1970) Nature 227, 680-685.
34. Towbin, H., Staehlin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.
35. Morrisey, J.H. (1981) Anal. Biochem. 117, 307-310.

Journal of General Virology (1992), 73, 639–651. Printed in Great Britain

Autoprocessing of the human immunodeficiency virus type 1 protease precursor expressed in *Escherichia coli* from a synthetic gene

Viviane Valverde,[1] Pierre Lemay,[1] Jean-Michel Masson,[1] Bernard Gay[2] and Pierre Boulanger[2]*

[1]Centre de Transfert en Biotechnologie-Microbiologie, CNRS-UA544, INSA, Complexe Scientifique de Rangueil, 31077 Toulouse and [2]Laboratoire de Virologie et Pathogénèse Moléculaires, Faculté de Médecine, 34060 Montpellier, France

A gene encoding an N-terminally extended precursor of 107 residues of the human immunodeficiency virus type 1 protease (PR107) was chemically synthesized and cloned into a bacterial expression vector, under the control of the *araB* promoter. PR107 was expressed alone or fused in phase to the amino or carboxy terminus of the bacterial β-galactosidase (β-gal). The yield of protease and β-gal was found to be significantly higher when the gene for PR107 was cloned upstream of the *Escherichia coli lacZ* gene (PR107–β-gal). Comparisons of the level of cloned protein expression between protease precursor and mature form suggested that this enhanced expression was due to the additional 5′ sequence of the PR107 gene, and occurred at the post-transcriptional level. Autoprocessing of protease precursor and its release from the β-gal fusion protein were analysed using wild-type and mutated cleavage sites. Mutations were introduced at amino acids downstream of the F–P scissile bond, at positions P4′ and P5′ in the C-terminal site (TLNF*PISP), and at position P3′ in a consensus N-terminal site (TLNF*PQITL) placed at the protease–β-gal junction. The data obtained suggested that (i) autoprocessing at the carboxy-terminal F–P bond was not significantly influenced by the presence of the N-terminal precursor sequence, (ii) P4′ and P5′ substitutions in the C-terminal site had no effect on cleavage, and (iii) P3′ in the N-terminal site tolerated a wide variety of substitutions.

Introduction

Specific processing of the human immunodeficiency virus type 1 (HIV-1) *gag–pol* polyprotein by the virus-encoded protease yields the structural *gag* gene products p24CA, p17MA and p15NC, as well as non-structural proteins, reverse transcriptase, endonuclease and protease (reviewed in Cann & Karn, 1989; Wills & Craven, 1991). The HIV-1 protease originates from a large *gag–prt–pol* polyprotein precursor, presumably as a result of intermolecular autoprocessing events (Lillehoj *et al.*, 1988; Mous *et al.*, 1988). It belongs to the aspartyl protease family and its active site contains the consensus D-T-G sequence (reviewed in Kraüsslich & Wimmer, 1988). It is active as a dimer (Katoh *et al.*, 1989; Miller *et al.*, 1989*b*) and is inhibited *in vitro* by pepstatin (Seelmeier *et al.*, 1988; Katoh *et al.*, 1987). Since specific proteolytic cleavages are essential for assembly of infectious HIV virions (Kohl *et al.*, 1988; Peng *et al.*, 1989; Gelderblom, 1991), the protease represents one of the possible targets for enzyme-directed anti-AIDS therapy (reviewed in Skalka, 1989; Tomasselli *et al.*, 1991). HIV-1 protease has been produced in bacteria (Debouck *et al.*, 1987; Graves *et al.*, 1988), chemically synthesized (Schneider & Kent, 1988), crystallized (Miller *et al.*, 1989*a*, 1990; Navia *et al.*, 1989) and co-crystallized with peptide-based specific inhibitors (Erickson *et al.*, 1990; Miller *et al.*, 1989*b*; Wlodawer *et al.*, 1989).

The aim of the present study was to analyse the mechanism of autocatalytic processing of HIV-1 protease at its N and C termini, and its subsequent release from a fusion protein. For this purpose, as an alternative to *in vitro* chemical or *in vivo* biological synthesis, the HIV protease was expressed from a synthetic gene cloned into a bacterial expression vector. We cloned the *prt* gene in *Escherichia coli*, unfused or fused in phase to the 5′ or 3′ end of the bacterial *lacZ* gene (upstream or downstream position, respectively). Our approach has several advantages compared to previously described production systems. (i) A synthetic gene sequence has more versatility for further genetic manipulations; (ii) the protease gene sequence was designed for optimal codon usage in *E. coli*; (iii) since the cloned protease was expected to be toxic for the recipient cell (Hostomsky *et al.*, 1989), the protease gene was cloned under the strong but tightly regulated *araB* promoter (Cagnon *et al.*, 1991); (iv) the yield of the β-galactosidase (β-gal)-fused gene products could be monitored easily using a simple β-gal enzymic assay; (v) purification of HIV-1 protease

0001-0519

might be achieved by affinity chromatography of protease–β-gal fusion on immobilized β-gal substrate or anti-β-gal antibody. Bacterial β-gal has also been recently used to monitor the HIV protease activity upon insertion of one of its specific sites into the *lacZ* gene (Baum *et al.*, 1990).

The HIV-1 protease sequence is flanked by two processing sites, SFNF*PQIT at its N terminus and TLNF*PISPI at its C terminus. Cleavage of the protease precursor and release from the protease–β-gal fusion protein was analysed by mutagenesis of amino acids downstream of the F–P scissile bond, at positions P4′ and P5′ in the C-terminal processing site and at position P3′ in a consensus N-terminal site placed at the protease–β-gal junction. The tolerance to amino acid residue substitutions was thus evaluated from the rate of protease–β-gal cleavage. We found that mutations at subsites P4′ and P5′ had no apparent effect on cleavage, but that the P3′ position in the N-terminal site showed some sensitivity to amino acid substitutions with strong effects on secondary structure.

Methods

Bacterial strains and plasmids. *E. coli* strain TG-1 (Amersham) was used for cloning, and MC1061 (Casadaban & Cohen, 1980) for gene expression. Cells were grown in LB medium supplemented with ampicillin (100 μg/ml; LBamp). The unfused protease gene was constructed from oligonucleotide blocks directly assembled into pCris10, a derivative of pKK233-2 (Amman & Brosius, 1985) which differs from pKK233-2 by its synthetic cloning cassette (V. Valverde *et al.*, unpublished results). β-gal–protease fusion constructs were directly assembled into pARA14. Plasmid pARA14 contains an arabinose-inducible promoter, termed *araB* (Cagnon *et al.*, 1991), and constituted our expression vector. For induction of gene expression, overnight cultures were diluted 20-fold in LBamp, and incubated at 37 °C until they reached an optical density at 600 nm (OD_{600}) of 0·5. Arabinose was then added to the cultures to a final concentration of 0·2%.

Nomenclature. PR100 and PR107 refer to the mature protease and protease precursor of 100 and 107 amino acid residues, respectively, including the initiator codon, methionine. When the *prt* gene was fused to the 5′ end of the *lacZ* gene (upstream position), the resulting in-phase fusion protein was called PR107–β-gal or PR100–β-gal. When *prt* was fused to the 3′ end of *lacZ* (downstream position) the name was β-gal–PR107. No β-gal–PR100 fusion protein was constructed since this protein would have lacked the protease N-terminal processing site at the β-gal–protease junction. The protease mutant named G33 had a glycine replacing the aspartyl residue 33 in the PR107 precursor, which corresponded to D (25) in the active site of the mature protease. The corresponding upstream fusion product was termed G33–β-gal. RR–β-gal designates a β-gal-fused protease variant in which the basic dipeptide Arg–Arg at positions P4′ and P5′ from the scissile bond at the PR107–β-gal junction was substituted for the neutral Pro–Gly dipeptide. When a sequence homologous to the protease N-terminal processing site, but containing a suppressible amber stop codon in lieu of the isoleucine codon at position P3′, was introduced at the downstream PR107–β-gal junction, the resulting plasmid was referred to as PR*amb*3′LZ. Amber suppressions at position P3′ in this junction sequence gave rise to P3′ mutants. A schematic drawing of our different gene constructs is presented in Fig. 1 (*a*).

Gene constructions. Oligonucleotides were synthesized on an Applied BioSystems 380A DNA synthesizer (Applied BioSystems) and purified as trityl derivatives on Nensorb Prep columns (NEN Research products) according to the manufacturer's specifications. They were enzymically phosphorylated as previously described (Normanly *et al.*, 1986). Two different HIV-1 *prt* genes were constructed. The first one encoded the mature form of protease of 100 amino acid residues, starting its sequence with the Met (1)–Pro (2) dipeptide whereas the second encoded a precursor form of 107 amino acids. The PR107 gene contained an additional 5′ nucleotide sequence encoding the heptapeptide (M)GTVSFNF (not counting the initiator methionine) present within the *gag–prt–pol* polyprotein, at the *gag*–protease junction, thus reconstituting the original Phe–Pro dipeptide of the N-terminal protease processing site (Darke *et al.*, 1988, 1989).

Both genes for the unfused protease constructs PR100 and PR107 were assembled stepwise from 20 complementary synthetic oligonucleotides annealed and ligated into three adapting building blocks (Fig. 1 *b*). Oligonucleotides corresponding to the protease coding strand were given odd numbers, and the even numbers corresponded to the non-coding strand (Fig. 1*c*). Step 1, oligonucleotides C1 to C8 were annealed and cloned between the *Hin*dIII and *Bgl*II restriction sites of pCris10. Step 2, the central part of the gene (oligonucleotides M1 to M6) was inserted between the *Bgl*II and *Nar*I sites. Step 3, the 5′ end of the PR100-encoding gene was reconstituted from oligonucleotide pairs N1-N2, N5-N6 and N7-N8, cloned between *Nar*I and *Nco*I; for the 5′ end of PR107, N3-N4 was inserted in lieu of N1-N2 (Fig. 1*c*, and inset *e*). For optimal expression, the sequence was designed according to the codon usage of *E. coli* (Fig. 1*c*). All gene constructs were verified by DNA sequencing (Sanger *et al.*, 1977). PR100 and PR107 genes were then inserted into the pARA14 expression vector, under the control of the arabinose-inducible promoter *araB* (Cagnon *et al.*, 1991).

In-phase β-gal fusions were directly constructed into the pARA14 expression vector already containing the gene for PR100 or PR107. The downstream fusion β-gal–PR107 was obtained by insertion of the plZ4 *Nco*I β-gal gene cassette (V. Valverde *et al.*, unpublished) into the *Nco*I site of PR107. The two upstream fusions PR100–β-gal and PR107–β-gal were generated by ligating the 5′ end of the *lacZ* gene to the 3′ end of the PR100 or PR107 gene. This was achieved by first cloning the L1 linker sequence (Fig. 1*d*) at the 3′ end of the *prt* gene. The plZ2 cassette (V. Valverde *et al.*, unpublished) was then inserted between the *Sal*I and *Hin*dIII sites, within the L1 and ARA14 sequences, respectively. The L1 linker was then mutated into the L0 linker (Fig. 1*d*), to obtain the PR100–β-gal and PR107–β-gal fusions.

Site-directed mutagenesis at the catalytic site and protease–β-gal junction. The phosphorothioate method (Taylor *et al.*, 1985) was used to introduce two substitutions into the active site of the protease, G and T substituting for A and C at nucleotides 100 and 101 of the PR107 gene, respectively (Fig. 1*d*). This created a new *Kpn*I site and changed the D (33) residue of the D-T-G triad (position 25 in the mature enzyme) into a glycine. This gave rise to the unfused G33 and β-gal-fused G33–β-gal mutant constructs (Fig. 1*a*).

The protease–β-gal junction in PR100–β-gal and PR107–β-gal fusions consisted of the sequence TLNF*PISP, representing the C-terminal processing site of the protease. By our cloning strategy (see above), the RR–β-gal fusion was obtained first, since it contained the L1 linker encoding the junction sequence LNF*PISRRGI. Using the same method (Taylor *et al.*, 1985), the L1 linker was then mutated into the L0 sequence (Fig. 1*d*), encoding the sequence LNF*PISPGGI. This latter sequence better mimicked the natural C-terminal processing site LNF*PISPIGI, in terms of side chain electric charges and overall conformation.

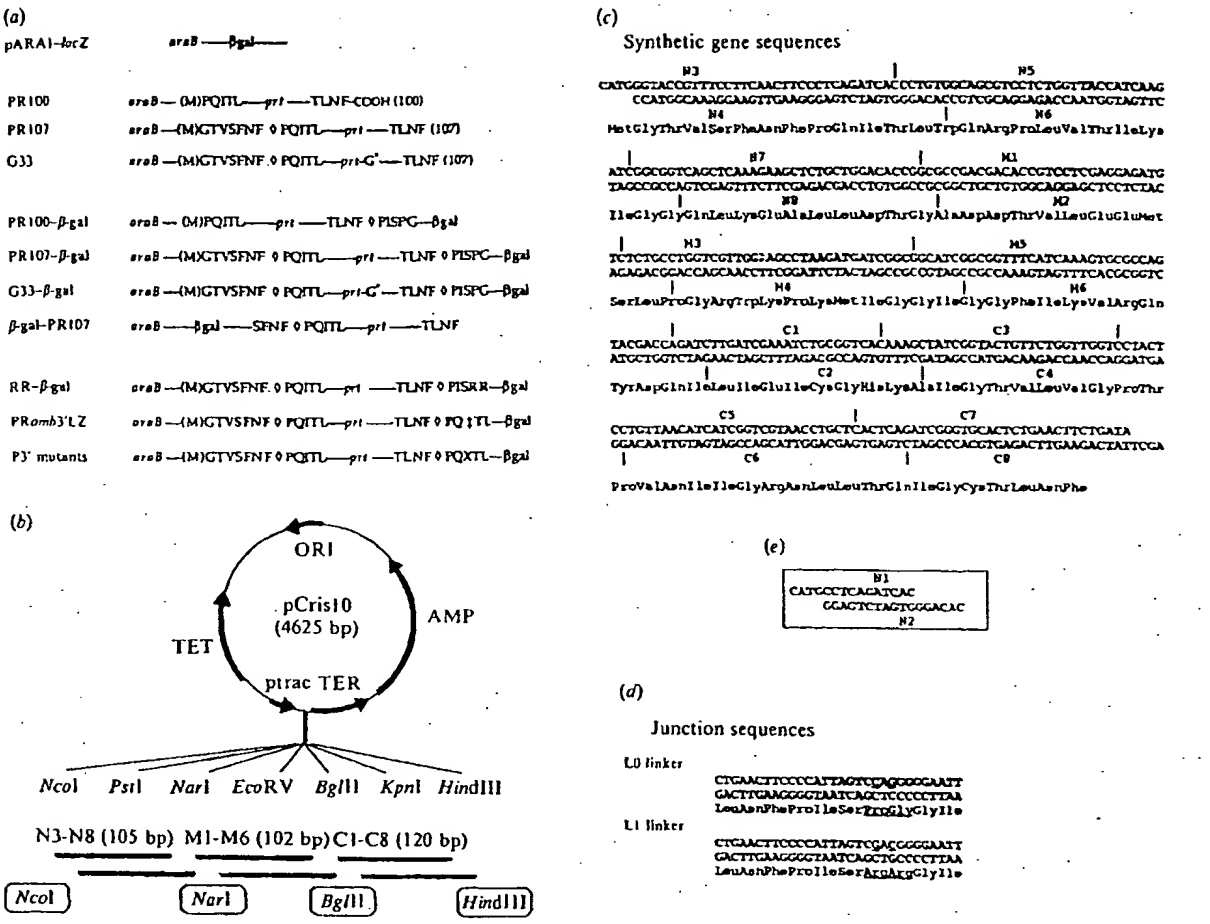


Fig. 1. (*a*) Schematic representation and nomenclature of the various protease gene (*prt*) constructs with the N- and C-terminal autoprocessing sites. The symbols and abbreviations used are: β-gal, β-galactosidase; *araB*, arabinose-inducible promoter; ◊, F–P scissile bond; ‡, amber mutation; G*, D to G substitution in the protease active site; X, amino acid specified by amber suppressor tRNA at the P3' position in the downstream processing site consensus to the N-terminal site. (*b*) Diagram of the stepwise construction of HIV-1 mature protease gene (PR100). Three successive building blocks with appropriate sticky ends were inserted into the pCris10 plasmid (Cagnon *et al.*, 1991). (*c*) Nucleotide sequence of the synthetic gene for the protease precursor of 107 residues (PR107), coding for the protease of 99 amino acids with an initiation methionine residue (PR100), and an additional N-terminal heptapeptide (GTVSFNF*) encoded by a portion of the *pol* sequence upstream of the *prt* gene. (*d*) Protease–β-gal junction sequences. L0 encodes the sequence LNF*PISPGGI, consensus to the C-terminal processing site; L1 for the mutant sequence LNF*PISRRGI, in which the proline–isoleucine dipeptide at the P4' and P5' subsites was replaced by Arg–Arg. In the inset (*e*) are shown the two complementary oligonucleotides used to construct the gene for PR100 protease, N1 and N2 replacing N3 and N4 at the 5', end.

Amber mutation suppressor tRNA scanning. A double-stranded oligonucleotide sequence (coding strand: 5' TGCACTCT-GAACTTCCCTCAGTAGACTGGGGATCCA 3') encoding the consensus N-terminal processing site TLNF*PQITL was synthesized, with an amber codon replacing the isoleucine codon at position P3' of the scissile bond. The TLNF*PQ*amb*TL-encoding nucleotide sequence was then inserted at the PR107–β-gal junction, between the *Apa*LI site at the 3' end of the precursor protease gene and the *Hin*dIII site in the pARA plasmid. The PR107 gene with the stop codon-containing junction sequence was then fused to the 5' end of the *lacZ* gene, giving rise to the PR*amb*3'LZ plasmid. A set of 12 amber tRNA suppressor genes cloned in pct2 vector (Kleina *et al.*, 1990*a*, *b*; Masson & Miller, 1986; Normanly *et al.*, 1986, 1990) were tested for co-expression with PR*amb*3'LZ in the MC1061 strain. The β-gal assay was used to estimate the efficiency of the amber suppression, as compared to the control PR107–β-gal fusion which contained a non-mutated C-terminal cleavage site as its junction sequence. Extracts from cells expressing PR*amb*3'LZ only were assayed for the β-gal assay and showed that there was no detectable reinitiation of protein synthesis on the *lacZ* mRNA downstream of the amber codon. The extent of protease autoprocessing in the presence of suppressor tRNAs was estimated by the protein ratio in the 127K and 116K bands (Fig. 3), representing the

protease–β-gal fusion protein and β-gal released from the protease, respectively. The 127K/116K doublets were scanned on anti-β-gal immunoblot patterns of SDS–polyacrylamide gels.

Enzyme assays. β-Gal activity was assayed in extracts from toluene-permeabilized cells by hydrolysis of *o*-nitrophenyl-β-D-galactoside (ONPG) (Miller, 1972). For protease assay, 1 ml aliquots of arabinose-induced cell culture taken at different time intervals were pelleted, washed in TN buffer (20 mM-Tris–HCl pH 7·5, 0·1 M-NaCl) and lysed in 50 µl of lysis buffer at 4 °C overnight with gentle mixing. Lysis buffer was made of 50 mM-Tris–HCl pH 7·5, 1 mM-DTT, 0·01% lysozyme, 0·1% NP40, 1 mg/ml DNase I and a cocktail of protease inhibitors (Darke *et al.*, 1988). The precursor to p24CA^{pro} produced in a baculovirus expression system (Royer *et al.*, 1991) was used as the substrate for protease. Aliquots (10 µl) of gag polyprotein (0·5 to 1·0 µg) in 0·2 M-sodium phosphate buffer pH 6·5, 1 M-NaCl, 5% glycerol and 0·25% NP40 were incubated with 10 µl of bacterial cell lysate at 27 °C and the reaction was stopped at different time intervals by addition of 20 µl of SDS–PAGE sample buffer. The gag precursor cleavage products were analyzed by SDS–PAGE and immunoblotting with gag-specific antibody.

Biochemical and immunological analyses. Bacterial proteins were analysed by SDS–PAGE (15%) in a discontinuous buffer system (Laemmli, 1970). In a typical experiment, 1 ml bacterial cell culture was pelleted, washed with TN buffer, resuspended in 100 µl of TN buffer, mixed with 100 µl of SDS–urea–2-mercaptoethanol sample buffer and denatured at 100 °C for 2 min. Gels were stained or electrically transferred to nitrocellulose membrane (BA85, Schleicher & Schüll) for 45 min at 180 mA in a semi-dry apparatus (Millipore SDE). Membranes were reacted with anti-β-gal (Research Plus) or anti-HIV-1 protease rabbit polyclonal antibody (a gift from Dr E. Cheng, DuPont) and a phosphatase-labelled anti-rabbit IgG conjugate (Sigma). Gag polyprotein products were detected on blots by reaction with an anti-pr55-p24CA^{pro} mouse monoclonal antibody (Epiclone-5001; Epitope Inc.) and a phosphatase-labelled anti-mouse IgG conjugate (Sigma). ELISA was performed using commercially available bacterial β-gal (Sigma) and anti-β-gal antibody (Research Plus).

For protein microsequencing, proteins were partially purified by acetone fractionation from bacterial cell lysates (Hansen *et al.*, 1988). The acetone precipitate was then chromatographed on FPLC-MonoS column in 50 mM-sodium 2-(*N*-morpholino)ethanesulphonic acid buffer pH 6·5, 1 mM-disodium EDTA, 10% glycerol (Pharmacia). Protease was eluted at 0·2 M-NaCl and its purification was achieved by preparative SDS–PAGE in a Tricine-buffered system (Schägger & von Jagow, 1987). For β-gal and protease, the 116K and 11K bands were transferred to Immobilon membranes (Ploug *et al.*, 1989) and amino-terminal sequences were determined using an Applied BioSystems 470A protein sequencer coupled to an Applied BioSystems 120A PTH Analyser.

Electron and immunoelectron microscopy. E. coli cells fixed in 2·5% glutaraldehyde in 0·1 M-phosphate buffer pH 7·5, for 1 h at 4 °C, were post-fixed with 2% osmium tetroxide in water and embedded in Epon (Epox-812; Fullam). Immunogold staining (IGS) was carried out by successive reactions of cell sections with primary anti-β-gal or anti-protease rabbit antibody overnight at 4 °C (at a dilution of 10 µg/ml in Tris-buffered saline), and 5 nm colloidal gold-labelled anti-rabbit IgG antibody for 1 h at room temperature (single IGS reaction). For double IGS, specimens were first simultaneously incubated with mouse monoclonal anti-β-gal antibody (Sigma) and rabbit polyclonal anti-protease antibody, then with 1 nm gold-labelled anti-rabbit IgG antibody and 5 nm gold-labelled anti-mouse IgG antibody (EM-GAM1 and EM-GAM5; BioCell Research Lab). Specimens were post-stained with 0·5% uranyl acetate and examined under the Philips EM-300 electron microscope.

Results

Expression of β-gal-fused and unfused HIV-1 protease and protease precursor in E. coli.

The level of expression of protease and β-gal synthesized in bacterial cells under the control of the *araB* promoter was estimated by SDS–PAGE and immunoblotting, and quantified by the β-gal assay. In *E. coli* expressing the unfused construct PR107, a band at 11K reacting with anti-protease antibody appeared after 1 h of arabinose induction (Fig. 2, lane 6) and progressively increased with the time of induction (Fig. 2, lanes 7 and 8). A similar pattern was obtained with the G33 mutant (Fig. 2, lanes 9 to 12), whereas no specific 11K band was detected in the PR100 pattern (Fig. 2, lanes 1 to 4). With the fused construct PR107–β-gal, the anti-protease serum revealed a discrete band of 127K fusion protein as early as after 0·5 to 1 h of induction (Fig. 3*a*, lanes 5 and 6) and an intense band at 11K became visible after 2 h (Fig. 3*a*, lanes 7 and 8). Several other discrete bands of intermediate cleavage products, ranging from 80K to 20K in apparent M_r were also seen on the blot (Fig. 3*a*, lanes 5 to 8). With the β-gal-fused mutant G33–β-gal, the 127K fusion band increased in intensity during the induction period, without detectable release of the 11K protease (Fig. 3*a*, lanes 13 to 16). In contrast to PR107, PR107–β-gal, G33 and G33–β-gal, a low level of expression of protease and fused β-gal–protease was obtained with PR100 and PR100–β-gal (Fig. 2, lanes 1 to 4; Fig. 4*a*, lanes 9 to 12), as well as with the downstream construct β-gal–PR107 (Fig. 3*a*, lanes 9 to 12). Immunoblotting with anti-β-gal antibody (Fig. 3*b*) confirmed the results obtained with the anti-protease serum (Fig. 3*a*). β-Gal was expressed at much higher levels with PR107–β-gal, G33–β-gal and RR–β-gal than with the other fusions β-gal–PR107 and PR100–β-gal (Fig. 3*b* and 4*b*).

The results of immunoblots were confirmed by β-gal assays performed on cell extracts: the level of β-gal expression with upstream fusions PR107–β-gal, G33–β-gal and RR–β-gal was ninefold higher than with the downstream fusion β-gal-PR107, and 15-fold higher than with PR100–β-gal (Table 1). The common feature between all the highly expressed genes was the presence of the additional 5′ sequence encoding the N-terminal precursor protease heptapeptide. To determine whether this enhanced expression of the cloned genes occurred at the transcriptional or post-transcriptional level, the overall quantity of protease mRNA expressed upon arabinose induction by the different fusion constructs was estimated by slot-blot hybridization, using ^{32}P-kinase-labelled oligonucleotide M6 as the ssDNA probe (Fig. 1*c*). No significant difference was found in the

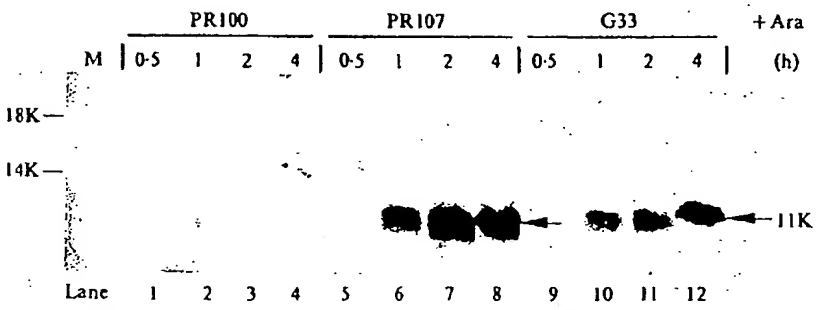


Fig. 2. SDS–PAGE and immunoblotting analysis of unfused HIV-1 protease expressed in *E. coli* under the control of *araB* promoter. Cells expressing PR100 (lanes 1 to 4), PR107 (5 to 8) and G33 mutant (9 to 12) were harvested after 0·5, 1, 2 and 4 h of arabinose induction, respectively. Lane M, prestained marker lane (18K and 14K markers are shown). Blot was reacted with anti-HIV-1 protease rabbit serum and phosphatase-labelled secondary antibody.

Table 1. *Enzymic assay of β-gal-fused protease yields by* E. coli *cells harbouring various expression vectors**

Vector	β-Gal activity (U/ml)	Ratio 1†	Ratio2‡
pARA1-*lacZ*	610	1·0	–
β-gal-PR107	595	0·9	–
PR100-β-gal	313	0·5	1·0
PR107-β-gal	4850	7·9	15·5
G33-β-gal	5110	8·4	16·3
RR-β-gal	4444	7·3	14·2

* Figures in the table are values of β-gal activity, determined by hydrolysis of ONPG and expressed in units per ml (U/ml) of permeabilized cell extracts (Miller, 1972). Each value represents the average of five individual results obtained in triplicate experiments (S.E.M., 5 to 10% of the activity values). The β-gal-specific activity was 1 x 10^5 U/mg, as estimated from stained gel scannings and ELISA. The value of 4444 to 4850 U/ml thus corresponded to 44 to 48 mg/l.

† Ratio of protease-fused β-gal to pARA1-*lacZ* value.

‡ Ratio of protease-fused β-gal to PR100-β-gal value.

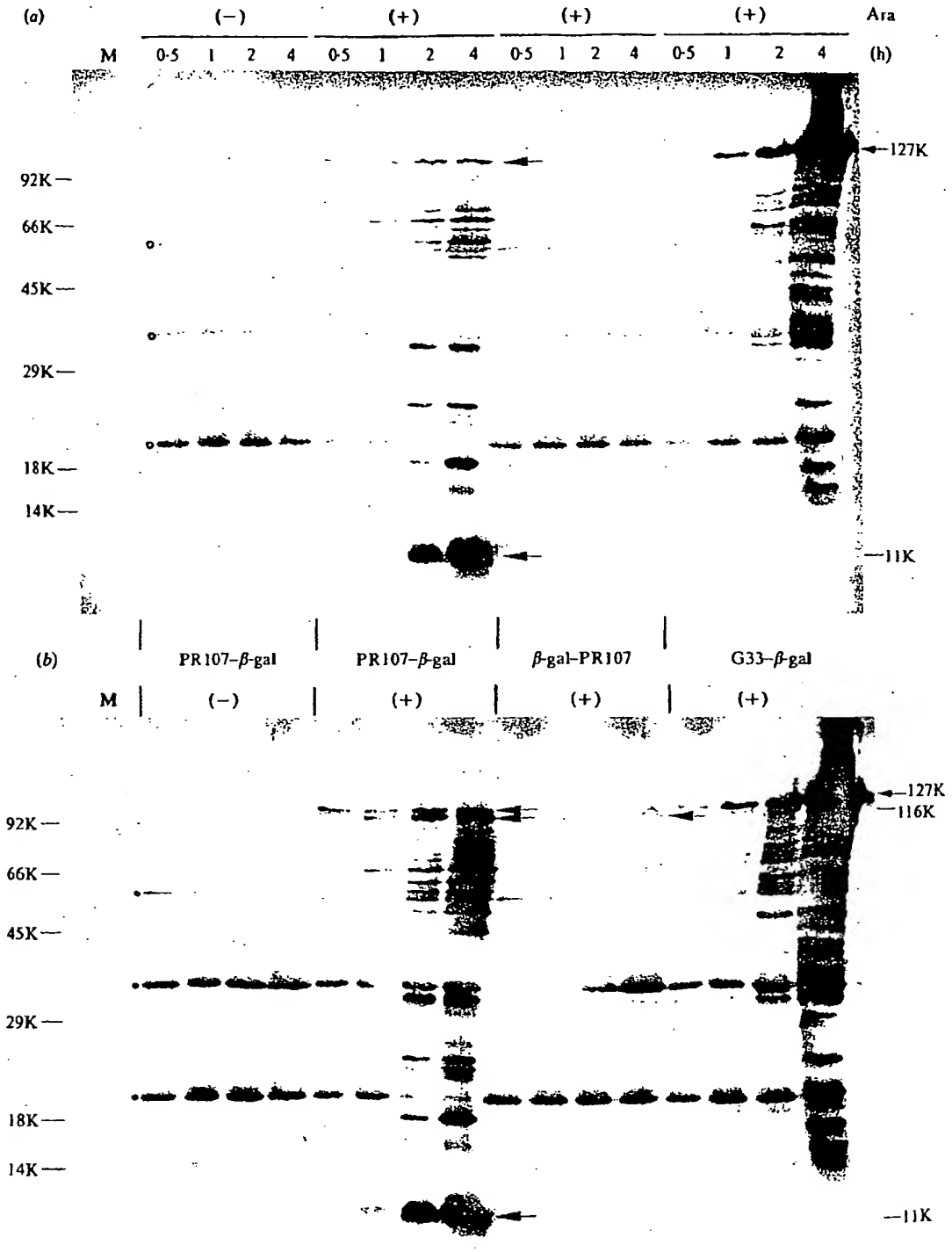
amount of protease mRNA expressed by the various constructs (data not shown), suggesting that the enhancing effect occurred at the post-transcriptional level.

The yield of HIV-1 protease obtained with the vector expressing the fusion construct PR107-β-gal was estimated by three indirect methods, all of them based on β-gal determination: (i) β-gal enzymic assays (Miller, 1972), from calculated specific activity of 10^5 units per mg of β-gal protein; (ii) ELISA, using commercial β-gal as the standard and anti-β-gal serum; (iii) Coomassie blue staining of the 127K/116K doublet band of protease-fused and released β-gal in SDS–PAGE, after calibrating the gel with purified β-gal. The protease (11K) represents about one-tenth of the mass of the fusion protein, in terms of polypeptide mass ratio (116K for β-gal). Thirty to 50 mg of PR107-β-gal fusion protein was obtained per litre of cell culture after 4 h of arabinose induction. Assuming a 100% efficiency for protease self-processing and release from the fusion protein, its theoretical recovery would be 3 to 5 mg of protease per litre, i.e. 1·2 to 2·0 mg per g of *E. coli* wet weight. A similar protease recovery (1 mg/l) was obtained from *E. coli* expressing a maltose-binding protein fusion construct (Louis *et al.*, 1991).

Activity, cytotoxicity and cellular distribution of β-gal-fused and unfused protease and protease precursor in E. coli

The results of the immunoblot analysis, showing evidence of HIV-1 protease release from the wild-type β-gal fusion constructs and accumulation of non-cleaved fusion protein with the G33-β-gal mutant (Fig. 3 and 4), suggested that the cloned protease was enzymically active and capable of self-processing *in vivo*. It could therefore be anticipated that protease synthesized in *E. coli* upon arabinose induction would also be active on its natural gag substrate. To verify its specificity of cleavage *in vitro*, our protease was incubated with baculovirus-expressed 41K gag polyprotein as a substrate ($pr41^{gag}$). $Pr41^{gag}$ comprised the p17MA and p25CA domains (amino acids 1 to 375 of the gag sequence), with only two unique protease sites (Mervis *et al.*, 1988), Tyr–Pro at the p17–p24 junction (amino acids 132 and 133), and Leu–Ala at the p24–p25 junction (positions 363 and 364), the Met–Met site at 377 and 378 being eliminated by carboxy truncation (Royer *et al.*, 1991).

As expected, the active site mutants G33 and G33-β-gal showed no detectable proteolysis of $pr41^{gag}$ (not shown), whereas all the other protease clones were found to cleave the gag polyprotein substrate at its two specific sites, yielding the characteristic anti-gag antibody-reacting p24–p25 doublet (Fig. 5). The highest activity was obtained with PR107 and PR107-β-gal, which converted almost all the gag precursor into $p24\text{-}p25CA^{gag}$ after 30 min of incubation (Fig. 5, lanes 3 and 7). The lower proteolytic activity shown by the extracts of cells



Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Fig. 3. SDS–PAGE and immunoblotting analysis of β-gal-fused protease clones PR107–β-gal (lanes 1 to 4 and 5 to 8), β-gal–PR107 (9 to 12) and G33–β-gal mutant (13 to 16), expressed in *E. coli* in the presence (+) or the absence (−) of arabinose inducer. The same blot was successively reacted with (*a*) anti-protease antibody and phosphatase-labelled conjugate, and (*b*) anti-β-gal antibody and its corresponding phosphatase-labelled conjugate. Cells were harvested after 0·5, 1, 2 and 4 h of induction. M, prestained *M*, markers (BRL, high *M*, range). Protease–β-gal fusion protein migrates as a 127K species, free β-gal and free protease as 116K and 11K protein bands, respectively.

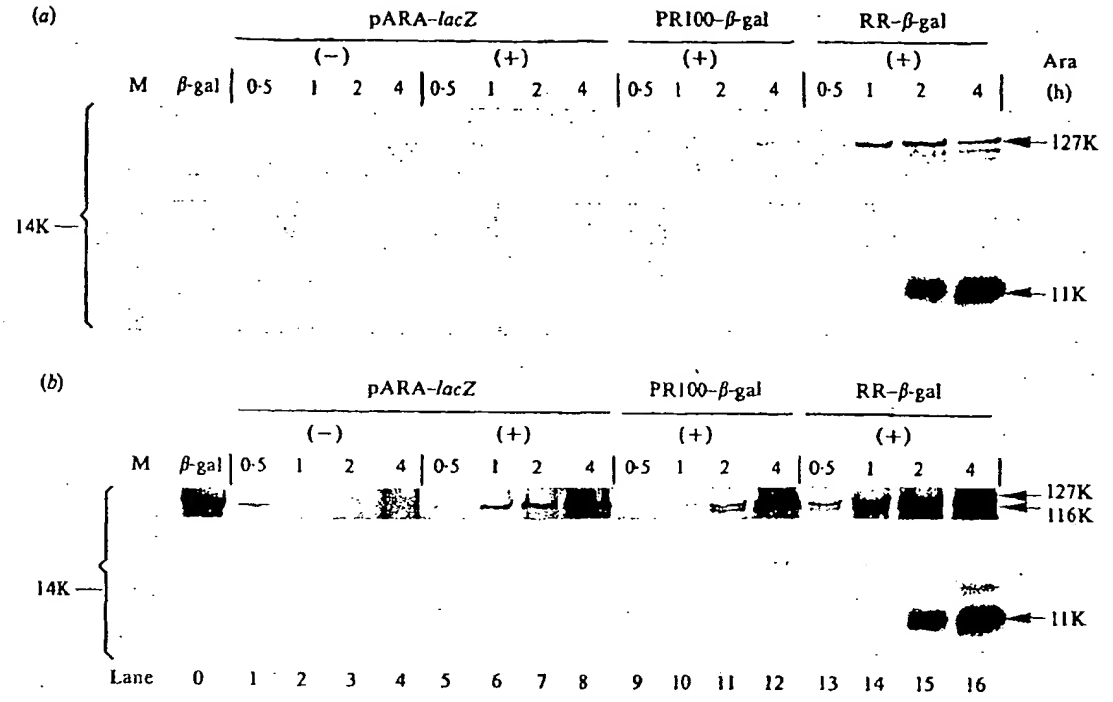


Fig. 4. SDS–PAGE and immunoblotting analysis of control pARA–*lacZ* (lanes 1 to 8) and β-gal-fused protease clones PR100–β-gal (lanes 9 to 12) and RR–β-gal mutant (13 to 16), expressed in *E. coli*. As in Fig. 3, the blot was successively reacted with (*a*) anti-protease antibody and phosphatase-labelled conjugate, and (*b*) anti-β-gal antibody and its corresponding phosphatase-labelled conjugate. Cells were maintained for 0·5, 1, 2 and 4 h in the presence (+) or absence (−) of arabinose. M, prestained M_r markers; β-gal, bacterial β-galactosidase (116K) from a commercial source. Note that the central portion of the blot, of no informative value, is not presented in this figure.

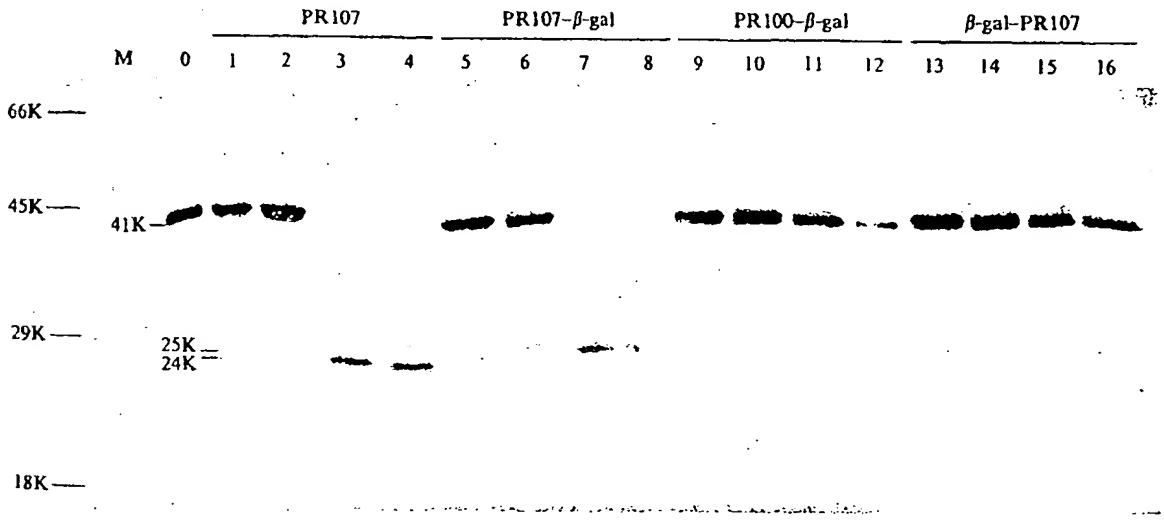


Fig. 5. Activity of bacterially expressed HIV-1 protease on its natural substrate *in vitro*. Extracts from *E. coli* cells harvested after 4 h of arabinose induction were incubated at 27 °C with baculovirus-expressed gag polyprotein pr41, precursor to p24–p25CA^{p9}. Samples were withdrawn at 5, 15, 30 and 60 min, and analysed in SDS–PAGE and immunoblotting with mouse monoclonal anti-gag precursor antibody (Epiclone 5001, Epitope) and phosphatase-labelled conjugate. Lane M, prestained M_r markers; lane 0, control gag polyprotein incubated for 60 min with extract from non-induced cells harbouring PR107.

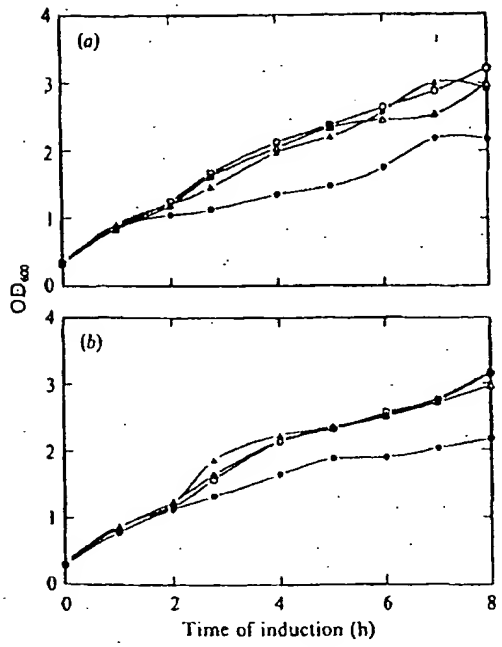


Fig. 6. HIV-1 protease expression and bacterial cell viability. Growth curves of *E. coli* cells harbouring various forms of β-gal-fused or unfused protease constructs were determined by measurements of OD_{600} throughout the arabinose induction period. (*a*) Non-induced (open symbols) or arabinose-induced (closed symbols) PR107 (circles) and PR107–β-gal (triangles). (*b*) Non-induced (open symbols) or arabinose-induced (closed symbols) mutants G33 (circles) and G33–β-gal (triangles).

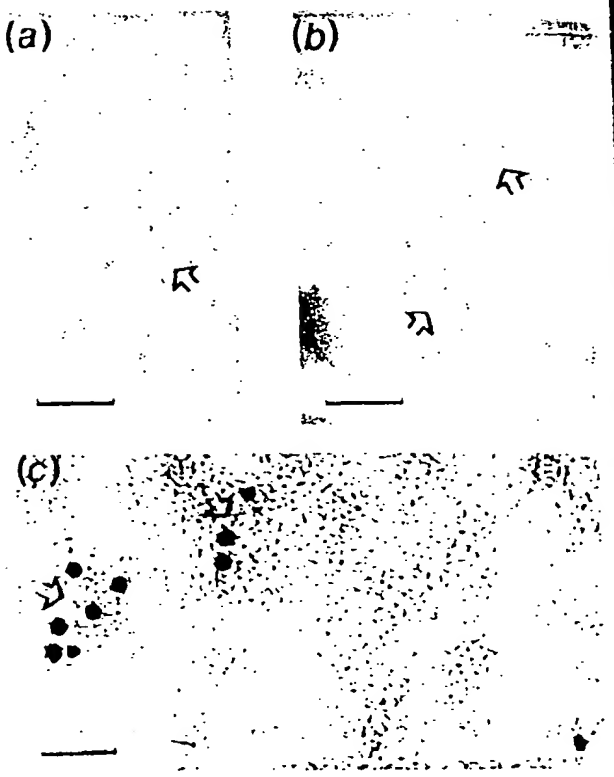


Fig. 7. Immunoelectron microscopy of PR107–β-gal fusion-expressing *E. coli* cells, harvested after 4 h of arabinose induction. (*a*) Single IGS with anti-β-gal rabbit antibody and 5 nm gold particle-labelled anti-rabbit IgG secondary antibody. (*b*) Single IGS with anti-protease rabbit antibody and 5 nm gold particle-labelled anti-rabbit IgG secondary antibody. The large open arrows point to antibody-reacting intracellular inclusion bodies. (*c*) Enlargement of an inclusion in cell doubly stained with primary anti-protease rabbit antibody and anti-β-gal mouse antibody, and 5 nm gold-labelled anti-mouse IgG (large open arrows) and 1 nm gold-labelled anti-rabbit IgG as the two secondary antibodies. Note that only a few 1 nm gold grains of the many more visible on the background are marked by thin arrows. Bar represents 200 nm in (*a*) and (*b*), and 20 nm in (*c*).

expressing PR100–β-gal and β-gal–PR107 (Fig. 5, lanes 9 to 12 and 13 to 16) reflected the lower levels of protease production from these two constructs (Fig. 2 to 4).

However, the difference in fused and unfused protease expression observed between some of our clones could be due to cytotoxicity of the protease and a negative feedback effect on protein synthesis. In addition, c.p.e. has been frequently found to be related to the cellular distribution of cloned foreign proteins. Arabinose-induced cells were therefore analysed with respect to their growth rate, and the cellular localization of unfused and β-gal-fused protease was examined by immuno-electron microscopy. As shown in Fig. 6, no detectable cytotoxicity was observed with the fusion constructs, and expression of only unfused PR107 and G33 proteases resulted in a slight reduction of the growth rate after 4 h of induction. The similar patterns shown by the clones expressing the active PR107 protease and the inactive G33 mutant suggested that the observed c.p.e. was due to the cellular accumulation of a foreign protein rather than to proteolytic activity of the protease *per se*.

When lysates from cells expressing unfused protease or β-gal alone were fractionated, most of the β-gal and protease activities were recovered in the soluble supernatant (not shown). No inclusion body was observed under the electron microscope, and IGS showed that the cloned proteins (protease or β-gal) were evenly distributed within the *E. coli* cytoplasm (not shown). In contrast, amorphous inclusions were found in cells expressing the two fused constructs PR107–β-gal and G33–β-gal. These intracellular inclusions reacted with both anti-β-gal and anti-protease antibodies in single IGS reactions (Fig. 7*a*, *b*). In the PR107–β-gal-expressing clone, the gold grain pattern was found to be different with anti-β-gal and anti-protease antibodies: β-gal molecules were observed inside and outside the inclusions, whereas protease was exclusively localized within the inclusion bodies (compare Fig. 7*a* and *b*). Double IGS, using 1 and 5 nm

colloidal gold particles specific for rabbit (anti-protease) and mouse (anti-β-gal) antibodies, respectively, confirmed that the inclusions contained both β-gal and protease molecules (Fig. 7*c*), and suggested that protease remained associated with the inclusion and in the vicinity of β-gal molecules after proteolytic processing.

Autoprocessing of HIV-1 protease: proteolytic cleavage at the protease–β-gal junction

For PR107–β-gal, as well as PR100–β-gal, cleavage at the F–P bond situated in the natural C-terminal site TLNF*PISP placed at the protease–β-gal junction was required to yield the mature protease (Fig. 1*a*). As already shown in Fig. 3 (*a*), the first band to appear at 0·5 h of induction in the PR107–β-gal pattern was a 127K fusion protein detected with anti-protease antibody. A protease band at 11K occurred in significant amounts after 2 h of induction, concurrently with a doublet at 127K and 116K (Fig. 3*b*). The 116K species was detected only with anti-β-gal antibody (Fig. 3*b*). No 11K band was visible in the G33–β-gal mutant although large quantities of fusion protein accumulated at 127K (Fig. 3*a*, *b*, lanes 15 and 16). All these data suggested that the protease was released from the β-gal fusion protein by a mechanism of specific self-processing at the C-terminal site of the protease and not by proteolysis due to bacterial proteases. Although produced in much lower yields, the 127K fusion band of PR100–β-gal was cleaved at the same time of induction as that of PR107–β-gal (Fig. 4), with the occurrence of detectable protease activity in the cell lysates (Fig. 5, lanes 11 and 12). This result suggested that, in the protease–β-gal fusion construct, cleavage at the C-terminal F–P bond of protease was not significantly influenced by the presence of cleaved or uncleaved N-terminal precursor sequence.

*Specificity of cleavage at the N-terminal SFNF*PQITL and C-terminal TLNF*PISP natural sites of the protease precursor*

Compared to mature protease PR100, PR107 precursor contained the additional N-terminal heptapeptide GTVSFNF present in the *pol* reading frame (Fig. 1*a*). To determine the specificity of cleavage at the N-terminal end of the protease precursor PR107 and at both the N and C termini of the β-gal-fused protease precursor PR107–β-gal expressed in *E. coli*, the 11K band produced by the PR107 clone and the 11K and 116K bands from the PR107–β-gal fusion construct were isolated and sequenced. The first six cycles showed the expected sequence P-Q-I-T-L-W for the two 11K proteases and P-I-S-P-G-G for the 116K processed β-gal. This confirmed that protease could mature from its PR107 precursor, by cleavage at the N-terminal site and the protease–β-gal junction, both cleavages occurring at specific F–P bonds.

Phenotype of mutants in the N- and C-terminal autoprocessing sites

Since any mutation upstream of the F–P scissile bond at the protease–β-gal junction or downstream of the F–P bond in the N-terminal site would alter the protease sequence and could therefore affect its enzymic activity, we restricted our mutational analysis to residues downstream of the C-terminal automaturation site. The P1′ and P2′ positions have been extensively studied (Margolin *et al.*, 1990; Partin *et al.*, 1990; Tritch *et al.*, 1991), whereas the requirements for the P3′, P4′ and P5′ subsites have not yet been clearly defined. The pattern of protease release by the RR–β-gal mutant, as shown in Fig. 4(*b*), was similar to its wild-type PR107–β-gal equivalent (Fig. 3*b*), although a slightly delayed processing (0·5 to 1 h) could be observed. This suggested that the presence of the two positively charged arginine residues at subsites P4′ and P5′ in the C-terminal processing site did not significantly impair the F–P scissile bond cleavage.

To analyse the influence of the amino acid residue at position P3′ in the N-terminal site on protease autoprocessing, the sequence TLNF*PQITL, consensus to the natural sequence SFNF*PQITL found at the N-terminus of the protease, was introduced at the PR107–β-gal junction. We constructed an appropriate vector, termed PR*amb*3′LZ, which expressed a β-gal-fused protease in which the protease sequence was flanked by two consensus N-terminal processing sites, i.e. its natural upstream SFNF*PQITL site and a downstream site with an amber stop codon in place of the isoleucine codon (TLNF*PQ*amb*TL; Fig. 1*a*). Any substitution at the P3′ subsite, downstream of the F–P bond, would therefore not affect the protease sequence. The *E. coli* strain harbouring PR*amb*3′LZ was then transformed by each of 12 plasmids expresssing an amber suppressor tRNA, and protease autoprocessing at the protease–β-gal junction was assayed by the occurrence of the 116K β-gal band released from the 127K protease–β-gal fusion in SDS–PAGE and immunoblotting.

As shown in Table 2, the efficacy of suppression, determined from the β-gal activity, was found to vary from 0·2% (lysine) to 41% (glycine) for the different suppressor tRNAs, a result which confirmed previous studies (Kleina *et al.*, 1990*b*). Protease autoprocessing was therefore analysed at late times of arabinose induction (4 h), to compensate for the low level of β-gal expression obtained with certain suppressors. Due to amino acid mischarging by certain suppressor tRNAs (Normanly *et al.*, 1990), some amino acid substitutions

Table 2. *Phenotype of amino acid substitutions at position P3' in the TLNF*PQambT autoprocessing site placed at the protease–β-gal junction**

Suppressor tRNA	β-gal activity (U/ml)	Suppression efficiency† (%)	Autoprocessing‡ (%)
PR107–β-gal alone	3631 (100%)	-	100
PR*amb3*'LZ alone	<2	0·0	-
PR*amb3*'LZ co-expressed with			
Gly2 (62% G; 37% E)	1489	41·0	70–75
HisA (94% H)	803	22·1	70–75
Thr2 (8% T; 86% K)	172	4·7	100
Ala2 (97% A)	159	4·3	100
GluA (59% E; 17% Q; 6% Y; 6%R)	129	3·5	100
Phe (98% F)	91	2·5	100
Arg (91% K; 5% R)	46	1·2	100
Pro (85% P)	38	1·0	60–70
Val (84% K; 5% V)	27	0·7	100
Cys (90% C)	23	0·6	100
Ile2 (93% K)	15	0·4	100
Lys (94% K)	7	0·2	100

* *E. coli* cells harbouring PR*amb3*'LZ were transformed with pct2 plasmid expressing one of the 12 suppressor tRNA genes listed in the table. The specificity and efficiency of the amino acid insertion by each suppressor is indicated in parentheses (Normanly *et al.*, 1990).

† The degree of amber suppression was estimated by β-gal activity, expressed as units per ml (U/ml; Miller, 1972). Controls were PR107–β-gal (100% control for β-gal activity and autoprocessing) and PR*amb3*'LZ alone (negative control for β-gal and suppression activity).

‡ Protease autoprocessing was estimated by scanning of the 127K/116K doublet bands of protease-fused and protease-released β-gal in immunoblots of SDS–polyacrylamide gels.

were under-represented (Table 2). Only three amino acids substituting for isoleucine at position P3' were found to have some discrete deleterious effect on the protease autoprocessing efficiency. Glycine, histidine (94% H inserted) and proline (85% P inserted) showed a slight but significant delay in protease processing. Gly2 tRNA inserted glycine and glutamic acid with similar efficiency, but comparison with GluA tRNA, which mainly inserted glutamic acid, allowed us to assign the observed effect on processing to the glycine substitution.

Discussion

In the present study we analysed HIV-1 protease autoprocessing using synthetic genes encoding a protease precursor of 107 amino acids (PR107) and the mature form of the enzyme (PR100), unfused and fused to the N terminus of β-gal (PR107–β-gal and PR100–β-gal), or to its C terminus (β-gal–PR107). It has to be kept in mind that the concept of 'autoprocessing' used for retroviral proteases has been based on the experimental observation that wild-type protease is released from protease-containing polyproteins expressed in *E. coli*, whereas the mutant protease domain is not (Debouck *et al.*, 1987, 1990; Kraüsslich & Wimmer, 1988; Loeb *et al.*, 1989; Mous *et al.*, 1988; Strickler *et al.*, 1989; and refer to Fig. 3). However, such results do not unambiguously prove that the PR domain embedded in a fusion construct is itself an active protease. It cannot be excluded that a low level of cellular proteolytic activity might randomly cleave the fusion protein and lead to release of a low amount of PR, which in turn acts in trans to generate further PR.

Yields of protease and β-gal were found to be significantly higher when the gene for PR107 was fused to the 5' end of the *E. coli lacZ* gene (PR107–β-gal), and comparison of the level of expression of unfused and β-gal-fused mature forms (PR100 and PR100–β-gal) suggested that the enhancing effect was post-transcriptional and due to the additional 5' sequence of the PR107 gene. In contrast to a previous report which showed an inactive form of β-gal-fused protease accumulating in inclusion bodies in *E. coli* (Giam & Boros, 1988), all our cloned proteases (except G33 and G33–β-gal protease mutants) were recovered in an active form under mild physiological conditions, and none of them, not even the inclusion-forming PR107–β-gal, required a solubilization step in urea-containing denaturing buffer. Our data are therefore reminiscent of previous observation on high protease yields obtained with certain clones of protease precursors encoding N-terminal sequences of 11 or 56 residues upstream of Pro (1) (Debouck *et al.*, 1990). According to our results, an upstream sequence of seven

residues is sufficient to enhance the protease production without altering protease activity.

Upstream fusion to β-gal, as in PR107–β-gal, did not result in any detectable cytotoxicity for *E. coli* cells (Fig. 6), and a comparison of PR107- and PR107–β-gal-expressing clones with the inactive mutants G33 and G33–β-gal suggested that the PR107- or G33-induced c.p.e. resulted from the expression of a foreign gene product in *E. coli* rather than to the enzymic properties of the protease itself. Only the two upstream fusion constructs PR107–β-gal and G33–β-gal gave rise to intracelluar inclusions. This is in contrast to previous reports in which intracellular precipitates were only observed with mature protease or with protease precursors containing both upstream and downstream sequences expressed in heat-shock response-deficient *E. coli* (Debouck *et al.*, 1990). Immunoelectron microscopy of the PR107–β-gal inclusions revealed that protease remained associated with the inclusion bodies, even after processing (Fig. 7), a phenomenon which could account for the absence of toxicity observed with our β-gal-fused protease gene constructs.

Processing at Phe–Pro bonds occurred *in vivo* with a similar efficiency at the PR107–β-gal and PR100–β-gal junctions, at the N terminus of the PR107 precursor and at the N and C termini of the β-gal-fused protease precursor PR107–β-gal. The N-terminal site was also correctly processed when placed at the protease C-terminal end. This suggests that each autoprocessing site contains its own information for cleavage, independent of protein context, and supports the hypothesis that protease autoprocessing occurs as a result of intermolecular interactions (Miller *et al.*, 1990). However, this does not imply that the N- and C-terminal processing are independent events, since the assay used cannot distinguish independence of these two reactions from a situation in which the C-terminal cleavage is dependent on the N-terminal cleavage, but the latter one occurs at a higher rate (Kraüsslich *et al.*, 1989; Strickler *et al.*, 1989). Moreover, a recent study using mutated fusion proteins expressed in *E. coli* indicated that altering one of the protease cleavage sites influences the cleavage at the non-mutated site (Louis *et al.*, 1991).

Upon substitution of the P–G dipeptide at positions P4′ and P5′ in the C-terminal site by the two positively charged amino acids R–R, protease release was only slightly delayed. This result was not surprising if one considers (i) the topography of residues P4′ and P5′, positioned downstream of the F–P scissile bond and outside of the active site cleft, (ii) the tolerance of an arginine residue at the P4′ position of the natural site at the $p9NC^{gag}$–$p6LI^{gag}$ junction (Debouck *et al.*, 1990) and (iii) the tolerance of an arginine residue at P1′ position in non-viral protein substrates of HIV-1 protease, e.g. troponin C and calmodulin (reviewed in Tomasselli *et al.*, 1991).

The upstream and downstream boundaries of HIV-1 protease constitute its two cleavage sites, SFNF*PQIT at its N terminus and TLNF*PISP at its C terminus, respectively. In the present study, we have focused on the subsite P3′ of the N-terminal autoprocessing site of the protease for the two following reasons. (i) It has been previously shown that cleavage at the F–P bond constituting the N terminus of the mature protease takes place significantly faster than cleavage at its C-terminal F–P bond (Strickler *et al.*, 1989), and that the natural N-terminal autoprocessing site VSFNF*PQITL had the highest V_{max}/K_m ratio (Kraüsslich *et al.*, 1989). (ii) As a result of mutational analyses of native substrates of HIV-1 protease (Partin *et al.*, 1990; Tomasselli *et al.*, 1991; Tritch *et al.*, 1991) and of its autoprocessing sites (LeGrice *et al.*, 1988; Loeb *et al.*, 1989; Louis *et al.*, 1991), and from theoretical considerations on the structure of the protease (Hellen *et al.*, 1989; Swanstrom *et al.*, 1989), the following consensus sequence for an autoprocessing site has been proposed: P4 (small and hydrophobic), P3 (undefined), P2 (small), P1 (aromatic or large and hydrophobic), P1′ (proline), P2′ (small and hydrophobic), P3′ (variable).

Amino acid residues at subsites P3 and P3′ have been postulated to be critical for the precise alignment of a peptide substrate within the protease active site cleft (Sali *et al.*, 1989; Miller *et al.*, 1990). In addition, once such a ligand is positioned within the cleft, P3 and P3′ are adjacent to both flaps of the protease dimer (Erickson *et al.*, 1990; Harte *et al.*, 1990; Lapatto *et al.*, 1989; Miller *et al.*, 1989*a*, *b*; Moore *et al.*, 1989; Navia *et al.*, 1989; Suguna *et al.*, 1987; Weber *et al.*, 1989; Wlodawer *et al.*, 1989). However, no experimental data are available on the influence of the residue at the P3′ subsite of the N-terminal site on protease autoprocessing, due to the fact that substitutions at this P3′ position would change the protease N-terminal sequence and therefore possibly alter its proteolytic activity. This was the case when aspartic acid at P3′ in the N-terminal site of the protease was substituted for isoleucine: no self-processing was observed with the mutant, whereas cleavage could be rescued in trans by the wild-type protease (Partin *et al.*, 1990).

For an indirect analysis of the effect of P3′ substitutions in the upstream site on protease autoprocessing, we introduced a consensus N-terminal processing site at the C-terminal extremity of the protease (at its junction with β-gal), and substituted a suppressible amber mutant codon for the isoleucine codon at position P3′ (TLNF*PQ*amb*T). When the amber mutation was assayed for rescue by a series of 12 suppressor tRNAs, only G, H and P substituting for I were found to reduce

the processing efficiency (Table 2). This suggested a tolerance for a wide variety of amino acid residues at P3′ of the N-terminal processing site of HIV-1 protease, except for amino acid substitutions with strong effects on the polypeptide chain secondary structure, e.g. glycine or proline. This implied a relatively high degree of flexibility of the protease flaps in their contact with the amino acid residue at P3′. The design of more efficient HIV protease inhibitors should take into account all available data on protease cleavage of viral and non-viral natural substrates (Rivière *et al.*, 1991), but the primary and therefore essential event is indeed its own cleavage and release by an autoprocessing mechanism.

The authors are grateful to Dr Edmund Cheng for providing them with anti-protease antibodies. This work has been supported by the Agence Nationale de Recherche sur le SIDA (ANRS, AC-2 and AC-14), the Conseil de la Région Midi-Pyrénées (JMM), the Conseil de la Région Languedoc-Roussillon (PB) and the Fondation pour la Recherche Médicale (PB). The original project was launched thanks to grant CRE 88-3001 from the Institut National de la Santé et de la Recherche Médicale.

References

AMMAN, E. & BROSIUS, J. (1985). 'ATG vectors' for regulated high-level expression of cloned genes in *Escherichia coli*. *Gene* **40**, 183–190.

BAUM, E. Z., BEBERNITZ, G. A. & GLUZMAN, Y. (1990). β-galactosidase containing a human immunodeficiency virus protease cleavage site is cleaved and inactivated by human immunodeficiency virus protease. *Proceedings of the National Academy of Sciences, U.S.A.* **87**, 10023–10027.

CAGNON, C., VALVERDE, V. & MASSON, J. M. (1991). *A* new family of sugar-inducible expression vectors for *E. coli*. *Protein Engineering* (in press).

CANN, A. J. & KARN, J. (1989). Molecular biology of HIV: new insights into the virus cycle. *AIDS* **3**, S19–S34.

CASADABAN, M. & COHEN, S. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular Biology* **138**, 179–207.

DARKE, P. L., LEU, C. T., DAVIS, L. J., HEIBACH, J. C., DIEHL, R. E., HILL, W. S., DIXON, R. A. F. & SIGAL, I. S. (1988). Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease. *Journal of Biological Chemistry* **264**, 2307–2312.

DARKE, P. L., NUTT, R. F., BRADY, S. F., GARSKY, V. M., CICCARONE, T. M., LEU, C. T., LUMMA, P. K., FREIDINGER, R. M., VEVER, D. F. & SIGAL, I. S. (1989). HIV-1 protease specificity of peptide is sufficient for processing of gag and pol polyproteins. *Biochemical and Biophysical Research Communications* **156**, 297–303.

DEBOUCK, C., GORNIAK, J. G., STRICKLER, J. E., MEEK, T. D., METCALF, B. W. & ROSENBERG, M. (1987). Human immunodeficiency protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. *Proceedings of the National Academy of Sciences, U.S.A.* **84**, 8903–8906.

DEBOUCK, C., DECKMAN, I. C., GRANT, S. K., CRAIG, R. J. & MOORE, M. L. (1990). The HIV-1 aspartyl protease: maturation and substrate specificity. In *Retroviral Proteases*, pp. 9–17. Edited by L. H. Pearl. London: Macmillan Press.

ERICKSON, J., NEIDHART, D. J., VAN DRIE, J., KEMPF, D. J., WANG, X. C., NORBECK, D. W., PLATTNER, J. J., RITTENHOUSE, J. W., TURON, M., WIDEBURG, N., KOHLBRENNER, W. E., SIMMER, R., HELFRICH, R., PAUL, D. A. & KNIGGE, M. (1990). Design, activity, and 2·8 Å crystal structure of a C_2 symmetric inhibitor complexed to HIV-1 protease. *Science* **249**, 527–533.

GELDERBLOM, H. R. (1991). Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS* **5**, 617–638.

GIAM, C. Z. & BOROS, I. (1988). *In vivo* and *in vitro* autoprocessing of human immunodeficiency virus protease expressed in *E. coli*. *Journal of Biological Chemistry* **263**, 14617–14620.

GRAVES, M. C., LIM, J. J., HEIMER, E. P. & KRAMER, R. A. (1988). An 11 kDa-form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity. *Proceedings of the National Academy of Sciences, U.S.A.* **85**, 2449–2453.

HANSEN, J. BILLICK, S., SCHULTZE, T., SUKROW, S. & MOELLING, K. (1988). Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibodies. *EMBO Journal* **7**, 1785–1791.

HARTE, W. E., SWAMINATHAN, S., MANSURI, M. M., MARTIN, J. C., ROSENBERG, I. E. & BEVERIDGE, D. L. (1990). Domain communication in dynamical structure of HIV-1 protease. *Proceedings of the National Academy of Sciences, U.S.A.* **87**, 8864–8868.

HELLEN, C. U. T., KRAÜSSLICH, H.-G. & WIMMER, E. (1989). Proteolytic processing of polyproteins in the replication of RNA viruses. *Biochemistry* **28**, 9881–9890.

HOSTOMSKY, Z., APPELT, K. & OGEDEN, R. (1989). High-level expression of self-processed protease in *Escherichia coli* using a synthetic gene. *Biochemical and Biophysical Research Communications* **161**, 1056–1063.

KLEINA, L. G. & MILLER, J. H. (1990*a*). Genetic studies of the *lac* repressor: extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *Journal of Molecular Biology* **212**, 295–318.

KLEINA, L. G., MASSON, J.-M., NORMANLY, J., ABELSON, J. & MILLER, J. H. (1990*b*). Construction of *E. coli* amber suppression genes: synthesis of additional tRNA genes and improvement of suppressor efficiency. *Journal of Molecular Biology* **213**, 705–717.

KATOH, I., YASUNAGA, Y. T., IKAWA, Y. & YOSHINAKA, Y. (1987). Inhibition of retroviral protease activity by an aspartyl protease inhibitor. *Nature, London* **329**, 654–656.

KATOH, I., IKAWA, Y. & YOSHINAKA, Y. (1989). Retroviral protease characterized as a dimeric aspartic proteinase. *Journal of Virology* **63**, 2226–2232.

KOHL, N. E., EMINI, E. A., SCHLEIF, W. A., DAVIS, L. J., HEIMBACH, J. C., DIXON, R. A. F., SCOLNICK, E. M. & SIGAL, I. S. (1988). Active immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences, U.S.A.* **85**, 4686–4690.

KRAÜSSLICH, H.-G. & WIMMER, E. (1988). Viral proteinases. *Annual Review of Biochemistry* **57**, 701–754.

KRAÜSSLICH, H.-G., INGRAHAM, R. H., SKOOG, M. T., WIMMER, E., PALLAI, P. V. & CARTER, C. A. (1989). Activity of purified biosynthetic proteinase of human immunodeficiency virus natural substrates and synthetic peptides. *Proceedings of the National Academy of Sciences, U.S.A.* **86**, 807–811.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 680–685.

LAPATTO, R., BLUNDELL, T., HEMMINGS, A., OVERINGTON, J., WILDERSPIN, A., WOOD, S., MERSON, J. R., WHITTLE, P. J., DANLEY, D. E., GEOGHEGAN, K. F., HAWRYLIC, S. J., LEE, S. E., SCHELD, K. G. & HOBART, P. M. (1989). X-ray analysis of HIV-1 proteinase at 2·7 Å resolution confirms structural homology among retroviral enzymes. *Nature, London* **342**, 299–302.

LEGRICE, S. F. J., MILLS, J. & MOUS, J. (1988). Active site mutagenesis of the AIDS virus protease and its alleviation by trans complementation. *EMBO Journal* **7**, 2547–2553.

LILLEHOJ, E. P., SALAZAR, F. H. R., MERVIS, R. J., RAUM, M. G., CHAN, H. W., AHMAD, N. & VENKATESAN, S. (1988). Purification and structural characterization of the putative gag-pol protease of human immunodeficiency virus. *Journal of Virology* **62**, 3053–3058.

LOEB, D. D., HUTCHISON, C. A., III, EDGELL, M. H., FARMERIE, W. G. & SWANSTROM, R. (1989). Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. *Journal of Virology* **63**, 111–121.

OLTS, J. M., MCDONALD, R. A., NASHED, N. T., WONDRAK, E. M., JERINA, D. M., OROSZLAN, S. & MORA, P. T. (1991). Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in *Escherichia coli*. *European Journal of Biochemistry* **199**, 361–369.

MARGOLIN, N., HEATH, W., OSBORNE, E., LAI, M. & VLAHOS, C. (1990). Substitutions at the P24 site of gag p17–p24 affects cleavage efficiency by HIV-1 protease. *Biochemical and Biophysical Research Communications* **167**, 554–560.

MASSON, J.-M. & MILLER, J. H. (1986). Expression of synthetic suppressor tRNA genes under the control of a synthetic promoter. *Gene* **47**, 179–183.

MERVIS, R. J., AHMAD, N., LILLEHOJ, E. P., RAUM, M. G., SALAZAR, F. H. R., CHANG, H. W. & VENKATESAN, S. (1988). The *gag* gene products of human immunodeficiency virus type 1: alignment within the *gag* open reading frame, identification of post-translational modifications, and evidence for alternative gag precursors. *Journal of Virology* **62**, 3993–4002.

MILLER, J. H. (1972). *Experiments in Molecular Genetics*. New York: Cold Spring Harbor Laboratory.

MILLER, M., JASKOLSKI, M., RAO, J. K. M., LEIS, J. & WLODAWER, A. (1989*a*). Crystal structure of retroviral protease proves relationship to aspartic protease family. *Nature, London* **227**, 576–579.

MILLER, M., SCHNEIDER, J., SATHYANARAYANA, B. K., TOTH, M. V., MARSHALL, G. R., CLAWSON, L., SELK, L., KENT, S. B. H. & WLODAWER, A. (1989*b*). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2·3 Å resolution. *Science* **246**, 1149–1152.

MILLER, M., SWAIN, A. L., JASKOLSKI, M., SATHYANARAYANA, B. K., MARSHALL, G. R., RICH, D., KENT, S. B. H. & WLODAWER, A. (1990). X-ray analysis of HIV-1 protease and its complexes with inhibitors. In *Retroviral Proteases*, pp. 93–106. Edited by L. H. Pearl. London: MacMillan Press.

MOORE, M. L., BRYAN, W. M., FAKHOURY, S. A., MAGAARD, V. W., HUFFMAN, W. F., DAYTON, B. D., MEEK, T. D., HYLAND, L., DREYER, G. B., METCALF, B. W., STRICKLER, J. E., GORNIAK, J. G. & DEBOUCK, C. (1989). Peptide substrates and inhibitors of the HIV-1 protease. *Biochemical and Biophysical Research Communications* **159**, 420–426.

MOUS, J., HEIMER, E. P. & LEGRICE, S. F. J. (1988). Processing protease and reverse transcriptase from human immunodeficiency virus type 1 polyprotein in *Escherichia coli*. *Journal of Virology* **62**, 1433–1436.

NAVIA, M. A., FITZGERALD, P. M. D., MCKEEVER, B. M., LEU, C.-T., HEIMBACH, J. C., HERBER, W. K., SIGAL, I. S., DARKE, P. L. & SPRINGER, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature, London* **337**, 615–620.

NORMANLY, J., MASSON, J.-M., KLEINA, L. G., ABELSON, J. & MILLER, J. H. (1986). Construction of two *E. coli* amber suppressor genes: $tRNA_{CUA}^{phe}$ and $tRNA_{CUA}^{Ser}$. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 6548–6552.

NORMANLY, J., KLEINA, L. G., MASSON, J.-M., ABELSON, J. & MILLER, J. H. (1990). Construction of *E. coli* amber suppressor genes: determination of tRNA specificity. *Journal of Molecular Biology* **213**, 719–726.

PARTIN, K., KRAÜSSLICH, H.-G., EHRLICH, L., WIMMER, E. & CARTER, C. (1990). Mutational analysis of a native substrate of the human immunodeficiency virus type 1 proteinase. *Journal of Virology* **64**, 3938–3947.

PENG, C., HO, B. K., CHANG, T. W. & CHANG, N. T. (1989). Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *Journal of Virology* **63**, 2550–2556.

PLOUG, M., JENSEN, A. L. & BARKHOLT, V. (1989). Determination of amino acid compositions and NH2 terminal sequences of peptides electroblotted onto PVDF membranes from tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis: application to peptide mapping of human complement component C3. *Analytical Biochemistry* **181**, 33–39.

RIVIÈRE, Y., BLANK, V., KOURILSKY, P. & ISRAËL, A. (1991). Processing of the precursor of NF-κB by the HIV-1 protease during acute infection. *Nature, London* **350**, 625–626.

ROYER, M., CERUTTI, M., GAY, B., HONG, S. S., DEVAUCHELLE, G. & BOULANGER, P. (1991). Functional domains of HIV-1 *gag*-polyprotein expressed in baculovirus-infected cells. *Virology* **184**, 417–422.

SALI, A., VEERAPANDIAN, B., COOPER, J. B., FOUNDLING, S. I., HOOVER, D. J. & BLUNDELL, T. L. (1989). High-resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: the analysis of the inhibitor binding and description of the rigid body shift in the enzyme. *EMBO Journal* **8**, 2179–2188.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SCHÄGGER, H. & VON JAGOW, G. (1987). Tricine sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry* **166**, 368–379.

SCHNEIDER, J. & KENT, S. B. H. (1988). Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. *Cell* **54**, 363–368.

SEELMEJER, S., SCHMIDT, H., TURK, V. & VON DER HELM, K. (1988). Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin. *Proceedings of the National Academy of Sciences, U.S.A.* **85**, 6612–6616.

SKALKA, A. M. (1989). Retroviral proteases: first glimpse at the anatomy of a processing machine. *Cell* **56**, 911–913.

STRICKLER, J. E., GORNIAK, J., DAYTON, B., MEEK, T., MOORE, M., MAGAARD, V., MALINOWSKI, J. & DEBOUCK, C. (1989). Characterization and autoprocessing of precursor and mature forms of human immunodeficiency virus type 1 (HIV 1) protease purified from *Escherichia coli*. *Proteins: Structure, Function and Genetics*, **6**, 139–154.

SUGUNA, K., PADLAN, E. A., SMITH, C. W., CARLSON, W. D. & DAVIES, D. B. (1987). Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus sinensis*: implications for a mechanism of action. *Proceedings of the National Academy of Sciences, U.S.A.* **84**, 7009–7013.

SWANSTROM, R., LOEB, D. D., EVERITT, L. & HUTCHISON, C. A., III (1989). Genetic analysis of HIV-1 protease. Viral proteinases as target for chemotherapy. In *Current Communications in Molecular Biology*, pp. 155–160. Edited by H. G. Kraüsslich, S. Oroszlan & E. Wimmer. New York: Cold Spring Harbor Laboratory Press.

TAYLOR, J. W., OTT, J. & ECKSTEIN, F. (1985). The rapid generation of oligonucleotide directed mutations at high frequency using phosphorothioate modified DNA. *Nucleic Acids Research* **13**, 8765–8785.

TOMASSELLI, A. G., HOWE, W. J., SAWYER, T. K., WLODAWER, A. & HEINRIKSOT, R. L. (1991). The complexities of AIDS: an assessment of the HIV protease as a therapeutic target. *Chimica oggi* **9**, 6–27.

TRITCH, R., CHENG, Y.-S. E., YIN, F. H. & ERICKSON-VIRTANEN, S. (1991). Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 *gag* polyprotein. *Journal of Virology* **65**, 922–930.

WEBER, I. T., MILLER, M., JASKOLSKI, M., LEIS, J., SKALKA, A. M. & WLODAWER, A. (1989). Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* **243**, 928–931.

WILLS, J. W. & CRAVEN, R. C. (1991). Form, function, and use of retroviral Gag proteins. *AIDS* **5**, 639–654.

WLODAWER, A., MILLER, M., JASKOLSKI, M., SATHYANARAYANA, B. K., BALDWIN, E., WEBER, I. T., SELK, L. M., CLAWSON, L., SCHNEIDER, J. & KENT, S. B. H. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245**, 616–621.

(*Received 30 July 1991; Accepted 28 October 1991*)

FEBS 18851

FEBS Letters 411 (1997) 231–235

The MBP fusion protein restores the activity of the first phosphatase domain of CD45

Hans K. Lorenzo[a], Donna Farber[1,b], Valérie Germain[b], Oreste Acuto[b], Pedro M. Alzari[a,*]

[a] *Unité d'Immunologie Structurale, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France*
[b] *Unité d'Immunologie Moleculaire, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France*

Received 14 April 1997; revised version received 28 May 1997

Abstract CD45 is a receptor-like protein tyrosine phosphatase critically involved in the regulation of initial effector functions in B- and T-cells. The protein comprises two phosphatase (PTP) domains in its cytoplasmic region. However, whether each PTP domain has enzyme activity by itself or whether both domains are required to build up a functional enzyme is unclear. We have studied different constructions of human CD45 comprising the two PTP domains, both separately and as a single protein, fused to maltose-binding protein (MBP). In apparent contrast with previous studies, we show that the first PTP domain of CD45 (when fused to MBP) may be a viable phosphatase in the absence of the second domain. Phosphatase activity resides in the monomeric form of the protein and is lost after proteolytic cleavage of the fusion partner, indicating that MBP specifically activates the first PTP domain. Furthermore, changes in the optimal pH for activity with respect to wild-type CD45 suggest that protein–protein interactions involving residues in the neighbourhood of the catalytic site mediate enzyme activation.

Key words: Receptor-like protein tyrosine phosphatase; Signal transduction; Maltose-binding protein; Fusion protein

1. Introduction

The activation and regulation of reaction cascades in signalling pathways leading to proliferation and differentiation in eukaryotic cells is related to the balance of tyrosine phosphorylation/dephosphorylation controlled by the action of tyrosine kinases and phosphatases (reviewed in [1,2]). Receptor-like protein tyrosine phosphatases (RPTP) exhibit a modular structure that includes one or two intracellular PTP domains, each homologous to soluble forms of monomeric phosphatases, a single membrane-spanning segment and a variable extracellular domain [3]. The PTP domain is organized as a central eight-stranded β-sheet flanked on both sides by α-helices, as revealed by the crystal structures of the soluble phosphatases PTP1B and Yop51 [4,5]. At the center of the active site, the highly conserved sequence motif (I/V)HCXAGXXR(S/T) contributes the essential nucleophilic cysteine residue and other functional groups required for phosphotyrosine binding and catalysis [6,7].

CD45, a prototypic receptor-like PTPase, is a 180–220 kDa protein highly expressed in hematopoietic cells. CD45 plays a critical role in the response of leukocytes to antigen, where it is involved in the early stages of the signal transduction pathways [8] (for review [9–11]). The two intracellular PTP domains of CD45 share about 40% of sequence identities with each other [3], and both contain a cysteine residue (Cys^{828} in PTP-I; Cys^{1144} in PTP-II) within the conserved sequence motif. However, several mutational studies have indicated that Cys^{828} (but not Cys^{1144}) is critical for phosphatase activity [12,13]. The substitution $Cys^{828} > Ser$ completely abrogated the enzymatic activity in recombinant forms [12–14] and in cells [15], but the replacement $Cys^{1144} > Ser$ in the second PTP domain resulted in a phosphatase with in vitro and in vivo properties similar to those of the wild-type enzyme [13,15]. These results suggest that only the first PTP domain, but not the second, behaves as an active phosphatase. In agreement with these observations, recombinant CD45–domain II alone yielded an inactive protein [13], and several RPTPs lack the conserved cysteine within the second PTP domain [16,17]. Furthermore, in RPTPs such as LAR and HPTPα, the first PTP domain alone displayed a phosphatase activity similar to that of the whole protein [18,19].

However, the first PTP of CD45 domain was found to be inactive when expressed independently [12,13], suggesting that in CD45, unlike other RPTPs, domain II is required for the activity of domain I. This hypothesis is further substantiated by the finding that a single point mutation in the second domain or the deletion of the region linking the two PTP domains totally abrogated the phosphatase activity of CD45 [20]. Contrasting these data, Tan et al. have shown in eukaryotic cells that the second domain of CD45 together with the C-terminal part of the first domain (without the catalytic cysteine) may also be a viable phosphatase [21]. Therefore, it is not clear at present whether the separate PTP domains of CD45 have phosphatase activity by themselves, or whether the enzymatic activity in one or both domains requires specific interdomain interactions.

We have studied different constructions containing the two phosphatase domains of CD45, both separately and as a single protein, fused to the maltose-binding protein (MBP). We found that the first PTP domain of CD45 when expressed as a MBP fusion protein is an active phosphatase in the absence of the second domain, and that the enzymatic activity is lost after proteolytic cleavage of the fusion partner, suggesting that specific interactions with an external factor are required to stabilize the active conformation.

2. Materials and methods

2.1. Materials

p-Nitrophenylphosphate (*p*-NPP), 2-β-mercaptoethanol (2-ME), dithiotreitol (DTT) and D-maltose were purchased from Sigma. Pefabloc® from Boehringer Mannheim. Factor Xa, antiserum anti-MBP and amylose resin were from New England BioLabs. Bradford reagent from Bio-Rads, and material for SDS-polyacrylamide gel elec-

*Corresponding author. Fax: (33) 1-45-68-86-07.
E-mail: alzari@pasteur.fr

[1] *Present adress:* Department of Microbiology, University of Maryland, 1109 Microbiology Building, College Park, MD 20742, USA.

0014-5793/97/$17.00
PII S0014-5793(97)00703-5

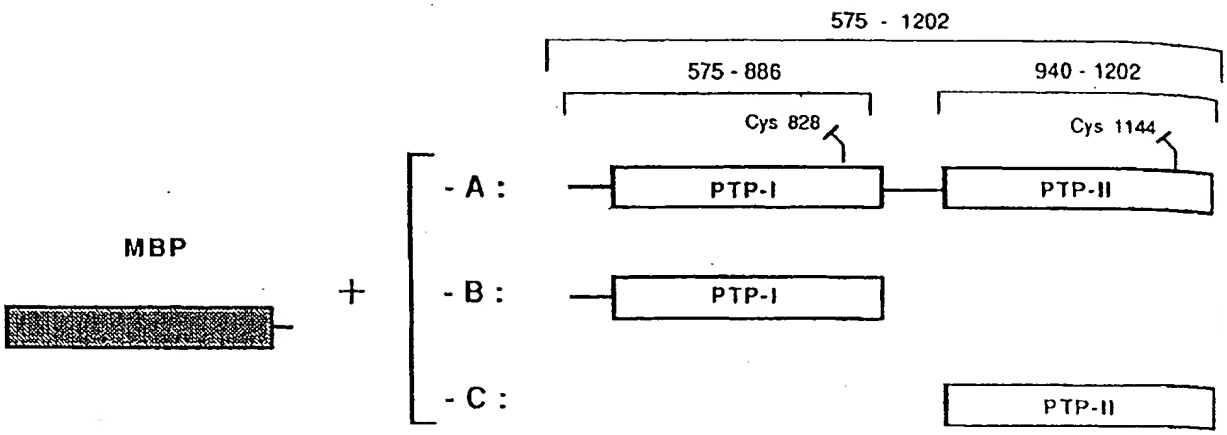


Fig. 1. Different constructions of human CD45. MBP: maltose-binding protein. Numbering of residues as in Streuli et al. (1987)[22].

trophoresis, gel filtration and ion exchange chromatography from Pharmacia.

2.2. MBP–CD45 fusion constructs

Three fusion constructs of human CD45 (PTP domain-I, PTP domain-II and PTP domain-I+II) were generated according to the boundaries proposed by Streuli et al. [22] (Fig. 1). Thus, MBP–PTP-I includes amino acid residues 575–886 (K_{575}IYDL...QALVE$_{886}$) of human CD45, MBP–PTP-II includes residues 940–1202 (Q_{940}EENKS... ...DVIAS$_{1202}$), and MBP–PTP-I+II includes residues 575–1202 (K_{575}IYDL... ...DVIAS$_{1202}$). The sequences were PCR amplified from an expression plasmid containing the entire human CD45 T200 coding sequence using the primers: 5'-GGAATTCAGATCTAAAATCTATGATCTACAT-3' and 5'-GGAATTCAGATCTTTACTGATTGTATTCCACCAA-3' (PTP-I); 5'-CGGGATCCCAAGAAGAAAATAAAAGT-3', and 5'-CGGGATCCTTAGCTGGCAATGACGTCATA-3' (PTP-II) and 5'-GGAATTCAGATCTAAAATCTATGATCTACAT-3' and 5'-GGAATTCAGATCTTTAGCTGGCAATGACGTCATA-3' (PTP-I+II).

Amplified PCR fragments were ligated to pPDXa [23] which had been previously linearized with *Bam*HI, treated with calf intestinal phosphatase, and gel purified. pPDXa contains the coding sequence for maltose binding protein (MBP) under the control of the maltose promoter and is a derivative of pMALcRI (New England BioLabs, Beverly, MA). All the plasmids were transformed into competent *malE*$^-$ PD28 cells according to established protocols [24].

2.3. Expression and purification of recombinant forms of human CD45

Cultures of PD28 cells, induction and isolation of fusion proteins were performed at 30°C using protocols previously described [25]. Protein purification was achieved by affinity chromatography in amylose resin as described by the furnisher, followed by ion exchange chromatography using a MonoQ column (Pharmacia) equilibrated in a buffer containing 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, and 2 mM DTT. The presence of different MBP–CD45 forms was evidenced by anti-MBP Western blot. Cleavage of purified fusion proteins was carried out by overnight incubation at 4°C with Factor Xa (1:100, w:w). The PTP domains were separated from MBP by anion exchange chromatography as described above. Protein purification was tested by SDS-PAGE and the respective concentrations measured by Bradford colorimetric assay.

2.4. PTPase assays

Kinetics of Michaelis-Menten were studied using *p*-NPP (10 mM final) as substrate in 50 mM imidazol (pH 7.0), 150 mM NaCl, 1 mM EDTA, and 0.1% 2-ME (total volume of 50 μl) at 37°C. The reaction was allowed to proceed for 5–20 min and quenched with 500 μl of 1 M NaOH. The amount of dephosphorylated substrate was calculated from the absorbance at 405 nm assuming a molar extinction coefficient $E_{405} = 18\,000\ M^{-1}\ cm^{-1}$ [6]. Michaelis-Menten constants were calculated using the non-linear regression program ENZFITTER [26].

The effect of pH on the PTPase activity of the recombinant forms of CD45 was studied using 10 mM *p*-NPP as substrate under similar conditions as above. Buffers used for these tests were as follows: pH 4–5.25, 100 mM acetate; pH 5.5–6.25, 100 mM citrate; pH 6.5–7.25, 100 mM imidazol; pH 7.5–8.5, 100 mM Tris-HCl. All buffers contained 1 mM EDTA, 0.1% 2-ME and 50 mM NaCl. pH dependence data were fitted by non-linear least-squares regression using the program KaleidaGraph 3.0.

3. Results and discussion

The cytoplasmic region of CD45 (PTP-I+II) as well as each separate PTP domain (PTP-I and PTP-II) (Fig. 1) were expressed as MBP fusion proteins in bacteria under control of the maltose promoter, and purified to near homogeneity (Fig. 2). Each construction was tested for PTPase activity using *p*-NPP as a substrate. Recombinant proteins containing the entire cytoplasmic region of CD45 (MBP–PTP-I+II and PTP-I+II) displayed a phosphatase activity (Table 1) comparable

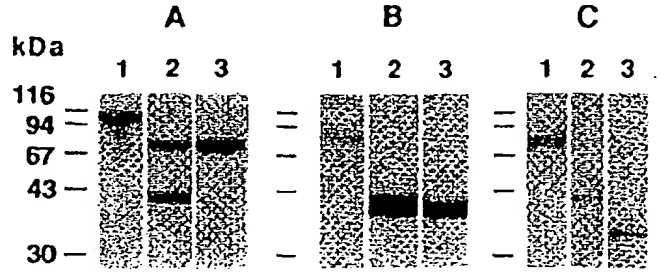


Fig. 2. SDS-PAGE analysis of MBP–CD45 constructions. A: MBP–PTP-I+II; B: MBP–PTP-I; C: MBP–PTP-II. Lane 1: after amylose affinity chromatography. Lane 2: after factor Xa treatment. Lane 3: after anion exchange chromatography.

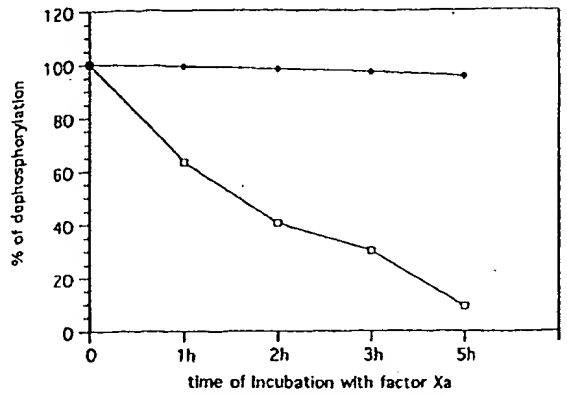


Fig. 3. Progressive loss of PTPase activity of MBP–PTP-I after treatment with factor Xa. MBP–PTP-I (□) and MBP–PTP-I+II (♦) were incubated with factor Xa at different times and their respective PTPase activities determined afterwards. pNPP (10 mM final) was used as substrate. Each point represents the mean value of quadruplicated experiences and was calculated as the percentage of dephosphorylation of pNPP exhibited. 100% of activity was considered as the activity displayed in identical conditions of incubation but in the absence of factor Xa.

to similar constructions studied by other authors [6,27,28]. Also, as expected, no enzymatic activity was detected in the recombinant proteins lacking the first phosphatase domain (MBP–PTP-II and PTP-II). However, MBP–PTP-I was found to be an active phosphatase, in contrast with previous work that reported the lack of catalytic activity of PTP-I in the absence of the second domain [12,13]. Both MBP–PTP-I and MBP–PTP-I+II displayed K_m values in the mM range, which are similar to the values reported by other authors for CD45 and other PTPases [6,27,28]. As shown in Table 1, V_{max} and k_{cat}/K_m values of MBP–PTP-I and MBP–PTP-II differ in about one order of magnitude, indicating that the absence of the second domain influences, but does not abrogate, the phosphatase activity of the first PTP domain. Similar results were obtained when two pY-peptides from hirudin and gastrin (Tyrosine Phosphatase Assay Kit, Boehringer) were used as substrates (data not shown).

Functional studies of different CD45 constructions suggested that specific interactions with the second domain are necessary to have a first active domain [13,20,21]. In a similar way, the phosphatase activity of MBP–PTP-I protein (in the absence of the second domain) might be accounted for either by specific interactions of PTP-I with MBP or by protein dimerization as observed, for example, in receptor-associated protein tyrosine kinases [29,30]. In order to test whether the covalently linked MBP was involved in PTP-I activation, MBP–PTP-I was treated with Factor Xa to separate the fusion partners. This experiment revealed a progressive loss of PTPase activity following the addition of Factor Xa (Fig. 3), indicating that the first PTP domain alone was not an active phosphatase. On the other hand, cleavage of the MBP moiety from the fusion protein containing the entire cytoplasmic region did not significantly affect the PTPase activity.

The possibility that the enzymatic activity of MBP–PTP-I activation could arise from protein dimerization was also investigated, since previous experimental evidences indicated that CD45 could be regulated by dimerization. For example, using cross-linking reagents in YAC-1 cell lysates, Takeda et al. detected CD45 homodimers apparently induced by a CD45-associated protein [31]. Furthermore, epidermal growth factor (EGF)-induced dimerization of an artificial EGF/CD45 chimera expressed in CD45-deficient T-cell line caused the loss of antigen-dependent activation [32]. Using gel filtration chromatography at different MBP–PTP-I concentrations, we observed monomers, dimers and higher oligomers of MBP–PTP-I. However, we have only found PTPase activity in fractions

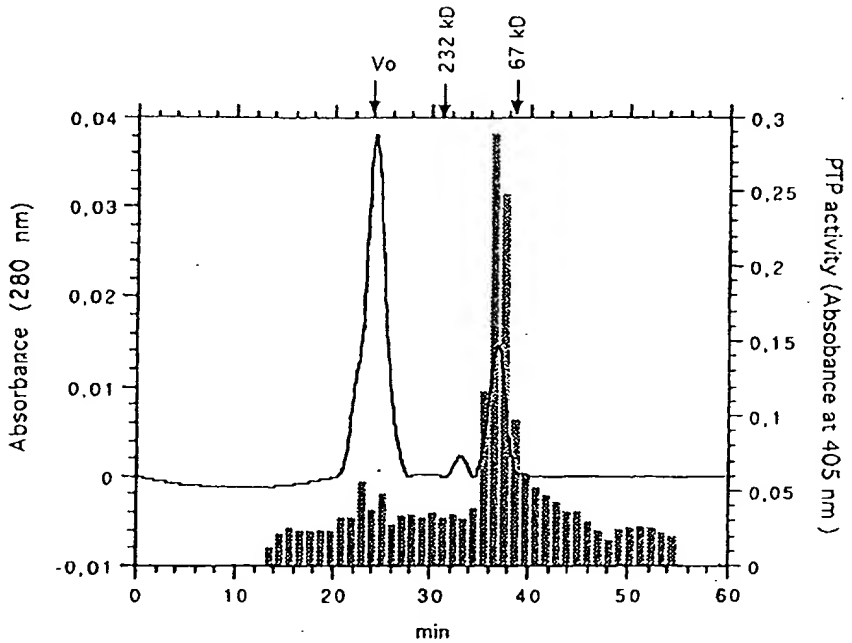


Fig. 4. The phosphatase activity is associated with the monomeric form of MBP–PTP-I after denaturation/renaturation. Aggregated (inactive) MBP–PTP-I was treated with 2.5% SDS and 5 M Urea and renatured by gel filtration (Superdex 200 SMART®). The fractions corresponding to the monomer, dimer and higher oligomeric forms were incubated with *p*-NPP (10 mM final) at 37°C for 12 h. The apparent PTP activity is represented by bars.

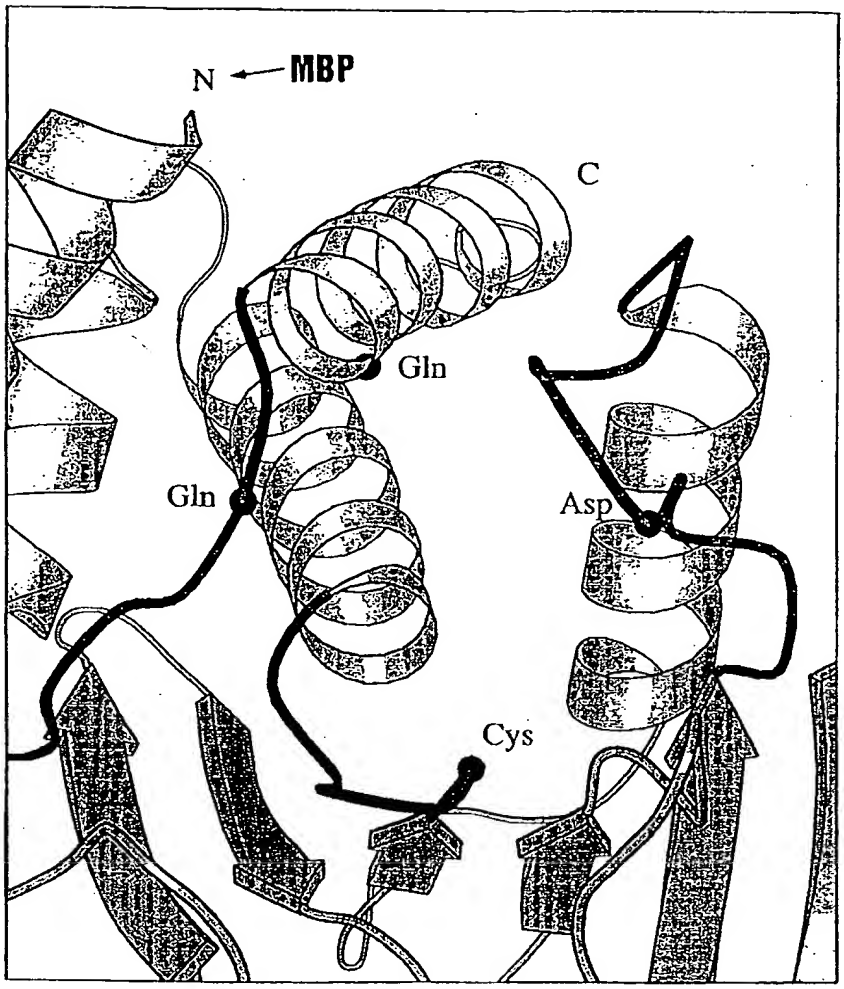


Fig. 5. Ribbon model of the active site of monomeric phosphatases. The nucleophilic cysteine residue is labelled. In our construction, the MBP moiety is fused to the NH_2 terminus of the PTP domain (indicated with an arrow) and could therefore interact with neighbouring PTP loops (shown in darker colour) containing functionally critical residues (Asp, Gln). The figure was drawn with MOLSCRIPT [36].

corresponding to the monomeric form, even after denaturation/renaturation of MBP–PTP-I (Fig. 4). These observations are consistent with the crystallographic study of RPTPα, which demonstrated that homodimerization of the first PTP domain inactivated the enzyme by blocking the access of substrate to the catalytic site [33].

At this point, our results demonstrated that: (i) the N-terminal PTP domain of CD45 can be an active phosphatase in the absence of the second domain, (ii) the PTP activity is only evidenced when PTP-I covalently bound to MBP (or to PTP-II in the wild-type enzyme) and (iii) the monomeric form of MBP–PTP-I is responsible for the phosphatase activity. In the light of these results, we may speculate that MBP activates the first PTP domain of CD45 through specific protein–protein interactions. This interaction probably involves contact residues close to the active site cleft and could compensate, at least partially, the putative contacts induced in wild-type CD45 by the presence of the second PTP domain. A structural model of the MBP–PTP-I protein provides additional support to this hypothesis. In the fusion protein, the C-terminal region of MBP is connected through a 22 amino acid-long linker to the N-terminal segment of PTP-I, a region which folds into an α-helix close to the substrate-binding cleft in monomeric phosphatases (Fig. 5). In particular, two loops of the PTP domains close to the N-terminal α-helix contain important functional residues. One of these loops has two glutamine residues (corresponding to Gln^{872} and Gln^{876} in CD45–PTP-I) which are highly conserved in the family of protein tyrosine

Table 1
Kinetic constants for dephosphorylation of *p*-NPP by different CD45 constructions

	K_m (mM)	V_{max} (μmol/min)	k_{cat}/K_m (s^{-1} M^{-1})	°C
MBP–PTP-I[a]	0.8 ± 0.2	22.98 ± 2.1	1233	37
MBP–PTP-I+II	1.75 ± 0.24	348.6 ± 1.2	12297	37
PTP-I+II[b]	4.8	36	11000	25

[a]Values were corrected assuming that only the monomeric form of MBP–PTP-I was an active PTPase.
[b]Values taken from Cho et al. (1993)[27].

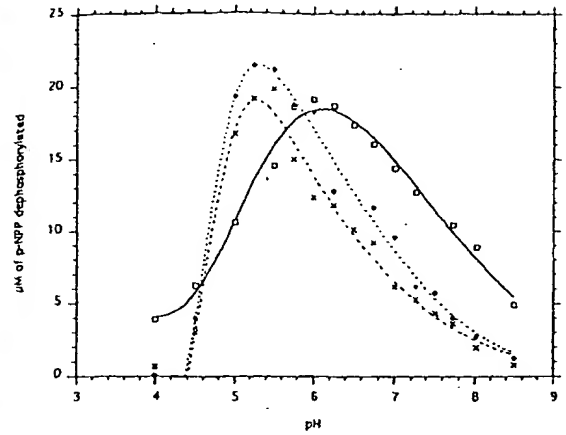


Fig. 6. Dephosphorylation of pNPP for the different active forms of CD45 as a function of pH. The solid line and (□) represents MBP–PTP-I, the dashed line and (X) represents PTP-I+II, and the dot line and (♦) represents MBP–PTP-I+II.

phosphatases and which have been proposed to be involved in substrate binding interactions [4,5]. A second loop close to the N-terminal α-helix of monomeric phosphatases contains the putative proton donor (Asp^{796} in CD45) and undergoes a significant conformational change upon the binding of phosphotyrosine to the active cleft (reviewed in [7]). Given their proximity to the N-terminal α-helix (and therefore to MBP), these loops might be involved in interdomain contacts, that could account for PTP-I activation.

Modifications of the local structural environment close to, or within, the active site cleft might affect the pKa value of functionally critical residues. The study of the pH dependence of the enzymatic activity could therefore provide an indirect evidence about MBP-contact regions in the neighbourhood of the PTP-I active site. As expected, the two constructions containing the entire cytoplasmic region (MBP–PTP-I+II and PTP-I+II) displayed similar bell-shaped profiles with a maximum at pH 5–5.2 (Fig. 6). These values are within the pKa ranges observed for other PTPases [6] and confirm our previous results (Fig. 3) showing that the presence of MBP did not affect the catalytic properties of PTP-I+II. However, the absence of the second PTP domain caused a displacement of the activity curve towards higher pH values, with an optimal activity at 5.8–6.2 for the MBP/PTP-I protein. These differences are apparently due to a change in the ascending slope of the activity curve and suggest that a region near the active site of the PTP-I domain might directly interact with the MBP moiety in the fusion protein. In the case of wild-type CD45, similar interactions between the two PTP domains could serve to regulate the PTPase activity [13,20,21], although other mechanisms of regulation, such as phosphorylation and external ligand binding have also been proposed [34,35]. Ultimate validation of this hypothesis can only be provided by further biochemical and structural studies of CD45 and other members of the RPTP family.

In conclusion, our results demonstrate that the first PTP domain of CD45 may be an active phosphatase in the absence of the second domain. However, a functional PTP-I domain requires specific protein–protein interactions with an additional factor (MBP or PTP-II), a particular feature of CD45 that may differentiate it from other receptor-like transmembrane phosphatases.

Acknowledgements: This work was supported by research grants from ARC (Association pour la Recherche sur le Cancer, grant no. 1141) and the Institut Pasteur. Hans K. Lorenzo was supported by a fellowship from the Colegio de España-Fundación Argentaria.

References

[1] Hunter, T. (1995) Cell 80, 225–236.
[2] Streuli, M. (1996) Curr. Op. Cell Biol. 8, 182–188.
[3] Charbonneau, H., Tonks, N.K., Walsh, K.A. and Fischer, E.H. (1988) Proc. Natl. Acad. Sci. USA 85, 7182–7186.
[4] Barford, D., Flint, A.J. and Tonks, N.K. (1994) Science 263, 1397–1404.
[5] Stuckey, J.A., Schubert, H.L., Fauman, E.B., Zhang, Z.Y., Dixon, J.E. and Saper, M.A. (1994) Nature 370, 571–575.
[6] Zhang, Z.Y., Wang, Y. and Dixon, J.E. (1994) Proc. Natl. Acad. Sci. USA 91, 1624–1627.
[7] Barford, D., Jia, Z. and Tonks, N.K. (1995) Nature Struct. Biol. 2, 1043–1053.
[8] Koretzky, G.A., Picus, J., Schultz, T. and Weiss, A. (1991) Proc. Natl. Acad. Sci. USA 88, 2037–2041.
[9] Chan, A.C., Desai, D.M. and Weiss, A. (1994) Annu. Rev. Immunol. 12, 555–592.
[10] Trowbridge, I.S. and Thomas, M.L. (1994) Ann. Rev. Immunol. 12, 85–116.
[11] Frearson, J.A. and Alexander, D.R. (1996) Immunol. Today 17, 385–391.
[12] Streuli, M., Krueger, N.X., Thai, T., Tang, M. and Saito, H. (1990) EMBO J. 9, 2399–2407.
[13] Johnson, P., Ostregaard, H.L., Wasden, C. and Trowbridge, I.S. (1992) J. Biol. Chem. 267, 8035–8041.
[14] Streuli, M., Krueger, N.X., Tsai, A.Y.M. and Saito, H. (1989) Proc. Natl. Acad. Sci. USA 86, 8698–8702.
[15] Desai, D.M., Sap, J., Silvennoinen, O., Schlessinger, J. and Weiss, A. (1994) EMBO J. 13, 4002–4010.
[16] Kaplan, R. et al. (1990) Proc. Natl. Acad. Sci. USA 87, 7000–7004.
[17] Krueger, N.X., Streuli, M. and Saito, H. (1990) EMBO J. 9, 3241–3252.
[18] Wang, Y. and Pallen, C.J. (1991) EMBO J. 10, 3231–3237.
[19] Cho, H., Ramer, S.E., Itoh, M., Kitas, E., Bannwarth, W., Burn, P., Saito, H. and Walsh, C.T. (1992) Biochemistry 31, 133–138.
[20] Ng, D.H.W., Mati, A. and Johnson, P. (1995) Biochem. Biophys. Res. Comm. 206, 302–309.
[21] Tan, X., Stover, D.R. and Walsh, K.A. (1993) J. Biol. Chem. 268, 6835–6838.
[22] Streuli, M., Hall, L.R., Saga, Y., Schlossman, S.F. and Saito, H. (1987) J. Exp. Med. 166, 1548–1566.
[23] Duplay, P., Thome, M., Hervé, F. and Acuto, O. (1994) J. Exp. Med. 179, 1163–1172.
[24] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
[25] Bedouelle, H. and Duplay, P. (1988) Eur. J. Biochem. 171, 541–549.
[26] Leatherbarrow, R.J. (1987) Enzfitter: A Non-linear Regression Data Analysis Program for the IBM PC, Elsevier, Amsterdam.
[27] Cho, H., Krishnaraj, R., Itoh, M., Kitas, E., Bannwarth, W., Saito, H. and Walsh, C.T. (1993) Prot. Sci. 2, 977–984.
[28] Pacitti, A., Panayiotis, S., Evans, M., Trowbridge, I. and Higgins, T.J. (1994) Biochim. Biophys. Acta 1222, 277–286.
[29] Heldin, C.H. (1995) Cell 80, 213–223.
[30] Johnson, L.N., Noble, M. and Owen, D.J. (1996) Cell 85, 149–158.
[31] Takeda, A., Wu, J.J. and Maizel, A.L. (1992) J. Biol. Chem. 267, 16651–16659.
[32] Desai, D.M., Sap, J., Schlessinger, J. and Weiss, A. (1993) Cell 73, 541–554.
[33] Bilwes, A.M., Hertog, J., Hunter, T. and Noel, J.P. (1996) Nature 382, 555–559.
[34] Stover, R. and Walsh, K.A. (1994) Mol. Cell. Biol. 14, 5523–5532.
[35] Autero, M. et al. (1994) Mol. Cell. Biol. 14, 1308–1321.
[36] Kraulis, J. (1991) J. Appl. Crystallogr. 24, 946–950.

JOURNAL OF VIROLOGY, Sept. 1985, p. 870–873
0022-538X/85/090870-04$02.00/0

Vol. 55, No. 3

Translational Readthrough of an Amber Termination Codon During Synthesis of Feline Leukemia Virus Protease

YOSHIYUKI YOSHINAKA, IYOKO KATOH, TERRY D. COPELAND, AND STEPHEN OROSZLAN*

Laboratory of Molecular Virology and Carcinogenesis, Litton Bionetics, Inc.,-Basic Research Program, National Cancer Institute-Frederick Cancer Research Facility, Frederick, Maryland 21701

Received 18 March 1985/Accepted 21 May 1985

Feline leukemia virus contains a protease which apparently has the same specificity as murine leukemia virus protease. It cleaves in vitro the $Pr65^{gag}$ of Gazdar-mouse sarcoma virus into the constituent p15, p12, p30, and p10 proteins. We purified the protease and determined its NH_2-terminal amino acid sequence (the first 15 residues). Alignment of this amino acid sequence with the nucleotide sequence (I. Laprevotte, A. Hampe, C. H. Sherr, and F. Galibert, J. Virol. 50:884–894, 1984) reveals that the protease is a viral-coded enzyme and is located at the 5′ end of the *pol* gene. As previously found for murine leukemia virus (Y. Yoshinaka, I. Katoh, T. D. Copeland, and S. Oroszlan, Proc. Natl. Acad. Sci. U.S.A. 82:1618–1622, 1985), feline leukemia virus protease is synthesized through in-frame suppression of the *gag* amber termination codon by insertion of a glutamine in the fifth position, and the first four amino acids are derived from the *gag* gene.

Feline leukemia virus (FeLV) is a non-genetically transmitted exogenous retrovirus shown to be associated with disease in domestic cats (4). As is characteristic of most retroviruses, FeLV genomic RNA contains three genes, *gag*, *pol*, and *env*, which are necessary for viral replication. The *gag* gene is translated into the polyprotein precursor $Pr65^{gag}$, which is processed to the structural proteins p15, p12, p30, and p10; the *pol* gene encodes an RNA-dependent DNA polymerase, and the *env* gene encodes the envelope glycoproteins of the virion surface (2). Based on the order of the *gag* gene-coded structural proteins, protein sequence homology, and immunological relatedness with murine leukemia virus (MuLV), FeLV is classified as type C, subgroup 1 (15). Further, Laprevotte et al. (10) reported a nucleotide sequence of 2,565 base pairs which includes a portion of the 5′ long terminal repeat, the *gag* leader, the complete *gag* gene, and 389 base pairs of the *pol* gene. Their data indicated that FeLV *gag* and *pol* genes are translated in different reading frames. Recently, we purified and sequenced Moloney (Mo)-MuLV protease responsible for the proteolytic processing of precursor polyprotein $Pr65^{gag}$ (18). The results showed that this protease is encoded by the *gag-pol* gene and synthesized within $Pr180^{gag\text{-}pol}$ through suppression of the amber termination codon located at the end of the *gag* gene.

In this report we describe the purification and partial sequence of a protease from FeLV, its location in the viral genome, and the translational control for its synthesis. FeLV (Rickard strain AB) was grown in feline lymphoblasts (16) and purified by sucrose density gradient centrifugation (13). In earlier studies we demonstrated Mo-MuLV protease activity under assaying conditions which involved endogenous substrate i.e., uncleaved $Pr65^{gag}$ of Mo-MuLV (19) released by disruption of the virus by Nonidet P-40. Subsequently, we adopted for routine analysis a method of assaying Mo-MuLV protease (20) with an exogenous substrate, Gazdar-mouse sarcoma virus (Gz-MSV) $Pr65^{gag}$ (6). Attempts to detect FeLV protease activity by the former method were unsuccessful because of the extremely low levels of uncleaved $Pr65^{gag}$ in purified FeLV. Therefore, in the present studies designed to purify FeLV proteolytic enzyme, we used Gz-MSV $Pr65^{gag}$ to assay FeLV protease activity. The suitability of Gz-MSV $Pr65^{gag}$ as a substrate was expected because the cleavage sites in the mouse and feline mature *gag* proteins are very similar based on protein and nucleotide seuqence data available for both systems (10, 11, 13, 17). The previously described methods for protease assay (18) and purification were used without major modifications.

To purified FeLV (158 mg) suspended in 2 ml of STE buffer (0.13 M NaCl, 0.01 M Tris hydrochloride [pH 7.2], and 0.001 M EDTA), 20 volumes of cold acetone (−70°C) was added, and the suspension was centrifuged at 4,000 × *g* for 10 min at 4°C. The precipitate was dried in vacuo. To solubilize the protease, the acetone powder was extracted (at 4°C for 30 min, stirred continuously) with 4 ml of TD buffer (0.02 M Tris hydrochloride [pH 7.0], 5 mM dithiothreitol [Sigma Chemical Co., St. Louis, Mo.]) containing 1.0 M NaCl. The extract was centrifuged at 20,000 × *g* for 20 min at 4°C. The supernatant was then fractionated on a Sephacryl S-200 column (2.5 by 90 cm) with TD buffer, and the protease activity was determined as described above by adding 100 μl of each fraction to 15 μg of Gz-MSV substrate in 1% Nonidet P-40. The protease-active fractions were pooled, lyophilized, and then further fractionated by reverse-phase high-pressure liquid chromatography (RP-HPLC) on a Bondapak C_{18} column (0.39 by 30 cm) (Waters Associates, Inc., Milford, Mass.). The protease activity was eluted with about 33% acetonitrile (Fig. 1A; fraction 30) and was detected by assaying lyophilized 5% aliquots of the fractions. When fractions 26 to 35 were incubated with disrupted Gz-MSV, fractions 27 to 33 cleaved $Pr65^{gag}$ into what appears to be the mature proteins p30, p15, p12, and p10 (Fig. 1B). The peak activity appeared in fraction 30. In addition, in fractions 28 and 32 intermediate cleavage products, presumably $Pr40^{gag}$ (p30 plus p10) and $Pr27^{gag}$ (p15 plus p12), were produced as observed in previous studies with Mo-MuLV protease (20). The purified protein, the majority of which eluted in fractions 29 to 31, showed a single band in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. From this analysis, the total protein was estimated to be approximately 7 μg.

* Corresponding author.

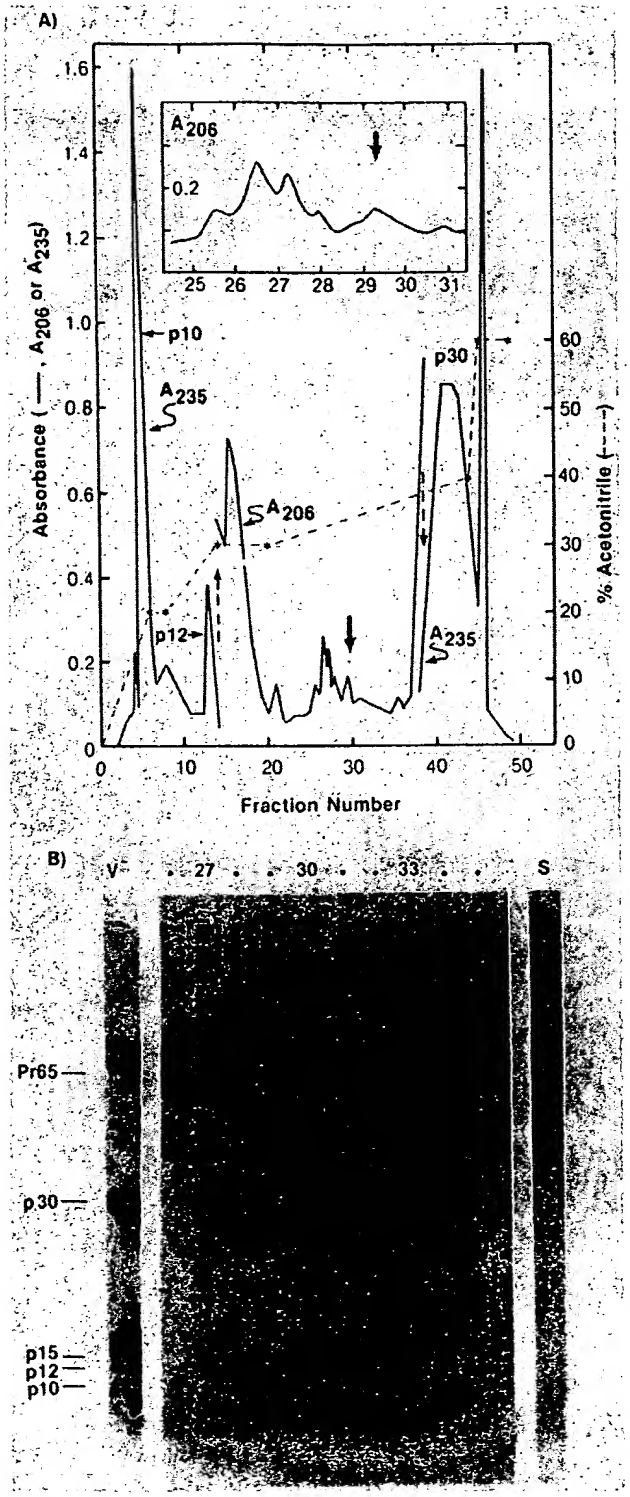


FIG. 1. Purification of protease by RP-HPLC. (A) Absorbance profile. Sephacryl S-200 chromatography fractions were applied to a Bondapak C_{18} column and then eluted with increasing concentration of acetonitrile as follows: 0 to 20% acetonitrile over 20 min, 20 to 30% acetonitrile over 30 min, and 30 to 60% acetonitrile over 90 min at a constant flow rate of 1.0 ml/min. Absorbance was measured at 235 nm or at 206 nm as indicated. (B) Assay of RP-HPLC fractions for protease activity. One-twentieth of each fraction was lyophilized and assayed for protease activity as described in the text. Proteins were visualized by staining with Coomassie brilliant blue R-250.

To determine the NH_2-terminal amino acid sequence of the protease, approximately 50 pmol of RP-HPLC-purified protein was subjected to automated Edman degradation in a gas-phase sequenator (9) with the program supplied by the manufacturer (Applied Biosystems, Inc). Conversion of the anilinothiazolinone amino acids to the phenylthiohydantoin amino acids was accomplished with 25% trifluoroacetic acid in water. Phenylthiohydantoin amino acids were identified and quantitated by RP-HPLC (8). The first 15 residues of the NH_2-terminal amino acid sequence were determined. The residue assignments together with the quantitative recoveries are given in Fig. 2A.

To examine whether the protease protein is viral coded, we aligned the experimentally determined sequence with the amino acid sequence deduced from DNA nucleotide sequence (10). The protease amino acid sequence begins with asparagine coded by triplet 2072 to 2074 and overlaps with the last four amino acids of the *gag* region (Fig. 2B). However, the third amino acid residue of protease was glycine and not glutamic acid, predicted from nucleotide sequence as shown in the alignment. The fifth amino acid, glutamine, corresponds to the *gag* termination codon TAG positioned at nucleotides 2084 to 2086. This is followed by a glutamic acid residue coded by the first triplet of the *pol* gene. The nucleotide sequence then continues in the *gag* reading frame and matches the protein sequence Thr-Gln-Gly-Gln-Asp-Pro-Pro-Pro-. At nucleotide 2113, however, we encounter a TGA stop codon in the DNA sequence. But if we remove one of the eight consecutive C residues from the sequence occupying positions 2104 through 2111, the nucleotide sequence matches both the FeLV protease sequence as determined here as well as the previously reported MuLV protease sequence (18) beyond nucloeotide 2111. The result is that FeLV protease is now in the same frame as *gag* and reverse transcriptase. This suggests that the DNA clone of FeLV strain B reported by Laprevotte et al. (10) is a noninfectious clone which, like many cloned DNAs of retroviruses, is defective. The protease NH_2-terminal amino acid sequence was found to be different from the nucleotide sequence of strain B at position 3 (Gly-Glu) and at position 7 (Thr-Ser). This may indicate strain differences and suggests that the major component of FeLV(AB) is strain A; the NH_2-terminal amino acid sequence analysis data of p10 (3) and p12 (unpublished data) also show differences from the strain B nucleotide sequence. In conclusion, these results show that the FeLV protease is a viral-coded enzyme and that it is synthesized by readthrough of the amber termination codon as in the murine system (18).

When we align the amino acid sequence of FeLV and MuLV protease deduced from the DNA sequence, 80% homology (25 different of 125 residues) is observed. Furthermore, the FeLV protease cleavage products of $Pr65^{gag}$ from Gz-MSV made sense when we compared the feline cleavage sites with the murine cleavage site between the *gag* proteins (p15, p12, p30, and p10); each feline cleavage site is very similar to the corresponding murine site (Fig. 2C). In both FeLV and Mo-MuLV systems (18), a single protease is responsible for the complete proteolytic processing of $Pr65^{gag}$ which is temporally linked to virus maturation. This conclusion regarding specificity is supported by the similarity in the chemical structure of all the cleavage sites (Fig. 2C). Although the peptide bonds cleaved (viz., tyrosyl-proline between p15 and p12, phenylalanyl-leucyl-proline between p12 and p30, leucyl-alanine between p30 and p10, and leucyl-asparagine-threonine between p10 and protease) are not identical, the carboxyl-terminal amino acid se-

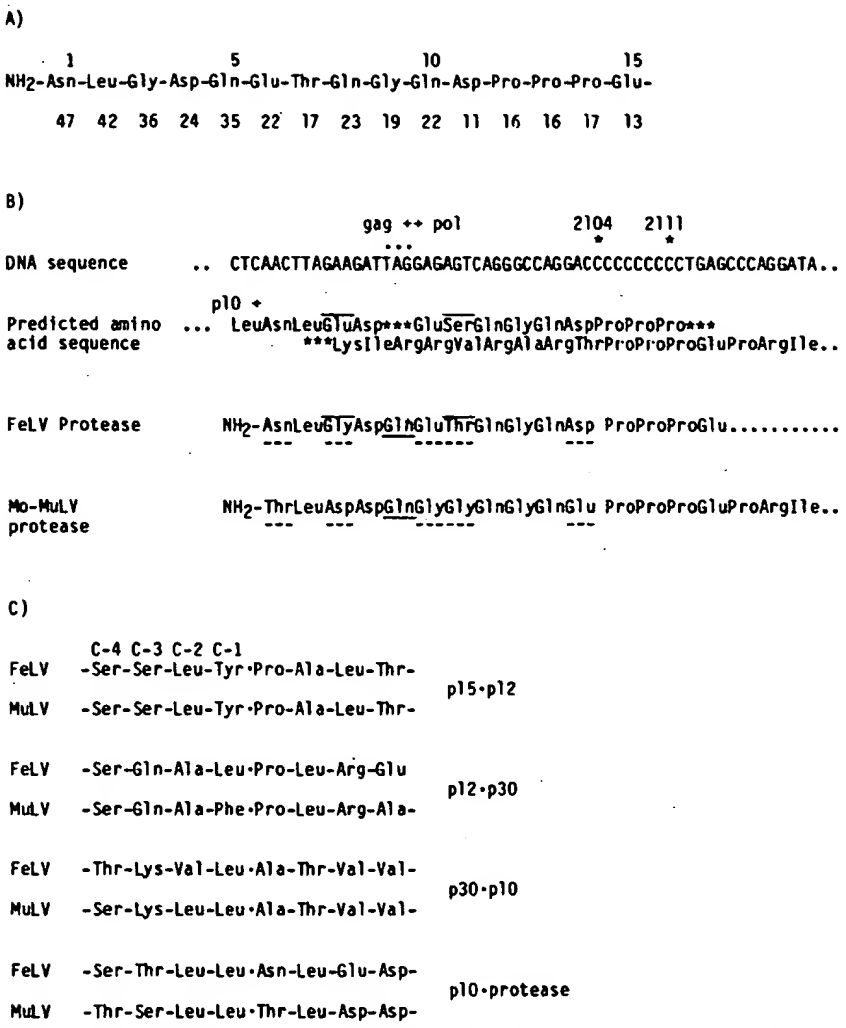


FIG. 2. (A) NH_2-terminal sequence of FeLV protease. Number below each residue is the yield (in picomoles) of the phenylthiohydantoin amino acid. (B) Alignment of the NH_2-terminal amino acid sequence with the DNA sequence of FeLV (10). The amber codon UAG is translated into glutamine (underline). Lines above amino acids indicate differences between deduced and determined sequences. Dotted lines under amino acids indicate differences between FeLV and Mo-MuLV protease (20). (C) Comparison of *gag* cleavage site sequences of FeLV and Mo-MuLV.

quences of the cleavage products are strikingly similar. They suggest a consensus (14) as follows: the penultimate residue (C-2) to the newly generated carboxyl terminus (C-1) is always a hydrophobic amino acid. The residue (C-3) next to it always has a polar side chain, charged or uncharged (serine, glutamine, and lysine), and the C-4 residue is either serine or threonine, both of which are known to initiate β-turns.

The exact mechanism of suppression is not clear. However, it is quite possible that the insertion of glutamine results from the misreading of the termination codon (UAG) by normal glutamyl tRNA as we proposed for Mo-MuLV (18). Other translational control mechanisms, such as suppression by nonsense suppressor tRNA (7), by splicing and by frame shift suppression (1, 5), have also been observed in both procaryotic and eucaryotic cell systems. The effects of the surrounding sequences on the suppression of a nonsense codon have also been investigated in bacteria (12). Retroviruses provide a useful model system for studying translational control in eucaryotic cells. It will be of interest to compare the influence of neighboring sequences on the suppression of certain amber termination codons in retroviral mRNAs.

We thank Young Kim for excellent technical assistance, and Jeannie Clarke and Sally Lockhart for preparation of the manuscript.

This research was sponsored by the National Cancer Institute, Department of Health and Human Services, under contract NO1-CO-23909 with Litton Bionetics, Inc., Kensington, Md.

LITERATURE CITED

1. **Atkins, J. F., and R. F. Gesteland.** 1983. Resolution of the discrepancy between a gene translation-termination codon and the deduced sequence for release of the encoded polypeptide. Eur. J. Biochem. 137:509–516.
2. **Baltimore, D.** 1975. Tumor viruses. Cold Spring Harbor Symp. Quant. Biol. 39:1187–1200.
3. **Copeland, T. D., M. A. Morgan, and S. Oroszlan.** 1984. Com-

plete amino acid sequence of the basic nucleic acid binding protein of feline leukemia virus. Virology 133:137–145.
4. **Essex, M.** 1980. Feline leukemia and sarcoma viruses, p. 205–229. *In* G. Klein (ed.), Viral oncology. Raven Press, New York.
5. **Fox, T. D., and B. Weiss-Brummer.** 1980. Leaky 1^+ and 1^- frameshift mutations at the same site in a yeast mitochondrial gene. Nature (London) 288:60–63.
6. **Gazdar, A. F., L. A. Phillips, P. S. Sarma, P. T. Peebles, and H. C. Chopra.** 1971. Presence of sarcoma genome in a "non-infectious" mammalian virus. Nature (London) New Biol. 324:69–72.
7. **Hatfield, D. L., B. S. Dudock, and F. C. Eden.** 1983. Characterization and nucleotide sequence of a chicken gene encoding an opal suppressor tRNA and its flanking DNA segments. Proc. Natl. Acad. Sci. U.S.A. 80:4940–4944.
8. **Henderson, L. E., T. D. Copeland, and S. Oroszlan.** 1980. Separation of all amino acids phenylthiohydantoins by high-performance liquid chromatography on phenylalkyl support. Anal. Biochem. 102:1–7.
9. **Hewick, R. M., M. M. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. 254:7990–7997.
10. **Laprevotte, I., A. Hampe, C. H. Sherr, and F. Galibert.** 1984. Nucleotide sequence of *gag* gene and *gag-pol* junction of feline leukemia virus. J. Virol. 50:884–894.
11. **Miller, A. D., and I. M. Verma.** 1984. Two base changes restore infectivity to a noninfectious molecular clone of Moloney murine leukemia virus (pMLV-1). J. Virol. 49:214–222.
12. **Miller, J. H., and A. Albertini.** 1983. Effects of surrounding sequence on the suppression of nonsense codons. J. Mol. Biol. 164:59–71.
13. **Morgan, M. A., T. D. Copeland, and S. Oroszlan.** 1983. Structural and antigenic analysis of the nucleic acid-binding proteins of bovine and feline leukemia viruses. J. Virol. 46:177–186.
14. **Oroszlan, S., and T. D. Copeland.** 1985. Primary structure and processing of *gag* and *env* gene products of human T-cell leukemia viruses $HTLV\text{-}1_{CR}$ and $HTLV\text{-}I_{ATK}$. Curr. Top. Microbiol. Immunol. 115:221–233.
15. **Oroszlan, S., and R. V. Gilden.** 1980. Primary structure analysis of retrovirus proteins, p. 299–344. *In* J. R. Stephenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
16. **Rickard, C. G., J. E. Post, F. Noronha, and L. M. Barr.** 1969. A transmissible virus-induced lymphocytic leukemia of the cat. J. Natl. Cancer Inst. 42:987–1014.
17. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) 293:543–548.
18. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. U.S.A. 82:1618–1622.
19. **Yoshinaka, Y., and R. B. Luftig.** 1977. Properties of a p70 proteolytic factor of murine leukemia viruses. Cell 12:709–719.
20. **Yoshinaka, Y., and R. B. Luftig.** 1982. P65 of Gazdar murine sarcoma viruses contains antigenic determinants from all four of the murine leukemia virus (MuLV) *gag* polyproteins (p15, p12, p30 and p10) and can be cleaved *in vitro* by the MuLV proteolytic activity. Virology 118:380–388.

JOURNAL OF VIROLOGY, Aug. 1994, p. 5045–5055
0022-538X/94/$04.00+0

Vol. 68, No. 8

Kinetic and Structural Analyses of Hepatitis C Virus Polyprotein Processing

RALF BARTENSCHLAGER,* LUDWINA AHLBORN-LAAKE, JAN MOUS, AND HELMUT JACOBSEN

Pharmaceutical Research-New Technologies, F. Hoffmann-La Roche Ltd., 4002 Basel, Switzerland

Received 9 February 1994/Accepted 22 April 1994

Recombinant vaccinia viruses were used to study the processing of hepatitis C virus (HCV) nonstructural polyprotein precursor. HCV-specific proteins and cleavage products were identified by size and by immunoprecipitation with region-specific antisera. A polyprotein beginning with 20 amino acids derived from the carboxy terminus of NS2 and ending with the NS5B stop codon (amino acids 1007 to 3011) was cleaved at the NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B sites, whereas a polyprotein in which the putative active site serine residue was replaced by an alanine remained unprocessed, demonstrating that the NS3-encoded serine-type proteinase is essential for cleavage at these sites. Processing of the NS3′-5B polyprotein was complex and occurred rapidly. Discrete polypeptide species corresponding to various processing intermediates were detected. With the exception of NS4AB-5A/NS5A, no clear precursor-product relationships were detected. Using double infection of cells with vaccinia virus recombinants expressing either a proteolytically inactive NS3′-5B polyprotein or an active NS3 proteinase, we found that cleavage at the NS4A/4B, NS4B/5A, and NS5A/5B sites could be mediated in *trans*. Absence of *trans* cleavage at the NS3/4A junction together with the finding that processing at this site was insensitive to dilution of the enzyme suggested that cleavage at this site is an intramolecular reaction. The *trans*-cleavage assay was also used to show that (i) the first 211 amino acids of NS3 were sufficient for processing at all *trans* sites and (ii) small deletions from the amino terminus of NS3 selectively affected cleavage at the NS4B/5A site, whereas more extensive deletions also decreased processing efficiencies at the other sites. Using a series of amino-terminally truncated substrate polyproteins in the *trans*-cleavage assay, we found that NS4A is essential for cleavage at the NS4B/5A site and that processing at this site could be restored by NS4A provided in *cis* (i.e., together with the substrate) or in *trans* (i.e., together with the proteinase). These results suggest that in addition to the NS3 proteinase, NS4A sequences play an important role in HCV polyprotein processing.

Infection with hepatitis C virus (HCV) is considered to be the major cause of posttransfusion and sporadic, community-acquired non-A, non-B hepatitis (for a recent review, see reference 23). It can lead to various clinical manifestations, including acute hepatitis, chronic hepatitis, liver cirrhosis, or an asymptomatic carrier state (22). In addition, the high prevalence of anti-HCV antibodies in chronically infected anti-hepatitis B antigen-negative patients with hepatocellular carcinoma indicates a strong association between HCV infection and tumor development (9, 12, 15, 29, 35).

Since the initial cloning (10), a number of new HCV isolates have been characterized (for a summary, see references 4 and 31). As deduced from infectivity studies with chimpanzees and from comparative sequence analyses, HCV has been classified as a separate genus in the family *Flaviviridae* together with pestiviruses and flaviviruses (7, 11, 27). These viruses have in common a virion with a lipid envelope, a single-stranded RNA genome of positive polarity encoding one long open reading frame of ca. 3,000 to 4,000 codons, and a similar genomic organization. The structural proteins (in the case of HCV, core-envelope 1 [E1]-E2) are located in the amino-terminal region and are followed by the nonstructural (NS) proteins (NS2, NS3, NS4A/B, and NS5A/B for HCV).

Production of mature proteins from the polyprotein precursor is accomplished by a series of proteolytic cleavages. Recent studies have shown that HCV structural proteins are generated from the polyprotein by host cell signalases (18, 19, 24, 26, 32), whereas processing of the NS polyprotein requires at least two virus-encoded enzymes: the NS2-3 proteinase (16, 20) cleaving between NS2 and NS3 and the NS3 proteinase cleaving at all the sites further downstream (2, 13, 17, 20, 36). Biochemical and mutational analyses suggest that the NS2-3 proteinase is a zinc-dependent metalloproteinase (20), which is, to our knowledge, without precedent among flaviviruses and pestiviruses. In contrast, NS3 appears to be a serine-type proteinase as are flavivirus NS3 and pestivirus p80 proteinases (5, 6, 8, 14, 28, 37).

In this report, recombinant vaccinia viruses were used to further characterize HCV polyprotein processing and to define the activity of the NS3 proteinase more precisely. Our results strongly suggest that the carboxy terminus of NS3 is generated by an intramolecular reaction, whereas cleavage at all sites further downstream can be mediated in *trans* by a proteinase containing the amino-terminal 211 amino acids of NS3. Furthermore, we have found that sequences from the NS4 region play an important role in polyprotein processing, especially for cleavage at the NS4B/5A site.

MATERIALS AND METHODS

Construction of plasmids for homologous recombination. The basic plasmids pATA 1007-2234, pATA 1007-1830, and pBSK 1007-1912 containing HCV sequences inserted into the modified vaccinia virus recombination vector pATA-18 or into the T7 transcription plasmid pBSK (Stratagene, Zürich, Switzerland) have been described before (2) (numbers refer to the first and the last amino acids of the expressed HCV polypro-

* Corresponding author. Present address: Institute for Virology, Johannes-Gutenberg University of Mainz, Hochhaus am Augustusplatz, 55101 Mainz, Germany. Phone: 49-6131-173134. Fax: 49-6131-172359.

tein fragment). To obtain plasmid pATA 1007-3011 bearing either a wild-type or mutated NS3 proteinase, pATA 1007-2234/wild type (wt) and pATA 1007-2234/S→A (in which the putative active site serine residue was replaced by an alanine residue [2]) were restricted with *Spe*I at the multiple cloning site (MCS) and *Hpa*I at HCV nucleotide position 5969 (according to the nomenclature of Kato et al. [25]) and combined with *Hpa*I-*Bsu*36I (5969 to 8134) and *Bsu*36I-*Spe*I (8134 to MCS) HCV fragments. These fragments were isolated from plasmids in which HCV sequences cloned from the serum of a chronically HCV-infected patient had been inserted (2). Plasmids containing HCV sequences from NS4A to NS5B (amino acids 1658 to 3011 of the polyprotein), NS4B to NS5B (1712 to 3011), or NS5A to NS5B (1973 to 3011) were constructed by PCR with upstream primers containing at their 5′ ends an *Eco*RI restriction site and an ATG codon. After 10 cycles, DNA fragments were purified by preparative gel electrophoresis, restricted with *Eco*RI (HCV nucleotide position 6687), and inserted into the *Eco*RI-digested plasmid pATA 1007-3011. Plasmid pATA 1007-1647 was obtained by insertion of an *Eco*RI (sticky end)-*Nsi*I (blunt ended after treatment with the Klenow enzyme) HCV fragment into pATA-18 cut with *Eco*RI (sticky end) and *Bam*HI (blunt ended). To construct plasmid pATA 1007-1564 an *Eco*RI (MCS)-*Stu*I (nucleotide position 5020) fragment isolated from pBSK 1007-1912 was inserted in pATA-18 restricted with *Eco*RI and *Spe*I. Plasmid pATA 1007-1355 was obtained by insertion of an *Eco*RI (MCS)-*Bst*EII (nucleotide position 4391) fragment into the *Eco*RI-*Sma*I-restricted pATA-18. To obtain plasmid pATA 1007-1395, an *Eco*RI-*Xba*I fragment (MCS to 4513) was isolated from plasmid pBSK Δ4513-5050 (2) and inserted into pATA-18 restricted with *Eco*RI and *Spe*I. Plasmids pATA 1007-1269 and pATA 1007-1238 were constructed in the same way with pBSK Δ4133-4783 and pBSK Δ4040-4713 (2) to isolate the HCV DNA fragments. Construction of the basic plasmids carrying various 5′-deleted NS3 fragments has been described recently (2).

Expression of HCV-specific proteins in *Escherichia coli* and generation of antisera. Expression of NS3-, NS3/4-, and NS5A-specific polyprotein fragments has been described previously (2) (see Fig. 2A). To express the NS5B-specific protein, a DNA fragment from nucleotide positions 7587 to 8207 (amino acids 2419 to 2625 of the polyprotein) was amplified by PCR with oligonucleotides carrying *Bam*HI restriction sites at their 5′ ends and was inserted into the *Bam*HI-restricted vector pDS 561/RBSII 6xHis (33). After induction, the protein was purified by metal chelate affinity chromatography under denaturing conditions (33). Further purification of the protein and generation of antisera were done exactly as described previously (2). The antibody titer as determined by enzyme-linked immunosorbent assay was 1:512,000.

Generation of recombinant vaccinia viruses. Human tk^- 143 cells were infected with the temperature-sensitive vaccinia virus *vts*-7 (kindly provided by Hans-Jürgen Schlicht, University of Ulm, Ulm, Germany) at a multiplicity of infection of 1 for 1 h at room temperature in Dulbecco's modified minimal essential medium (DMEM). After removal of the inoculum, cells were incubated in medium supplemented with 10% fetal calf serum (FCS) for 2 h at 33°C. Medium was removed, cells were washed several times with phosphate-buffered saline (PBS), and the calcium phosphate precipitate, prepared as described previously (3) with 1 μg of vaccinia virus wild-type DNA and 0.5 to 2 μg of plasmid DNA in a total volume of 250 μl, was added dropwise. After 1 h at room temperature, DMEM without FCS was added, cells were incubated for 2 h at 40°C and washed four times with PBS, and FCS-containing medium was added. Cells were lysed after 48 h at 40°C in 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid, and the lysate was used for infection of human tk^- 143 cells for 1 h at room temperature as described above. Finally, complete medium containing 100 μg of 5-bromo-2′-deoxyuridine (Sigma, Deisenhofen, Germany) per ml was added, and cells were incubated for 48 h at 37°C. Large-scale preparations of recombinant vaccinia viruses were grown on HeLa cells, and titers of infectious progeny were determined by plaque assay on human tk^- 143 cells.

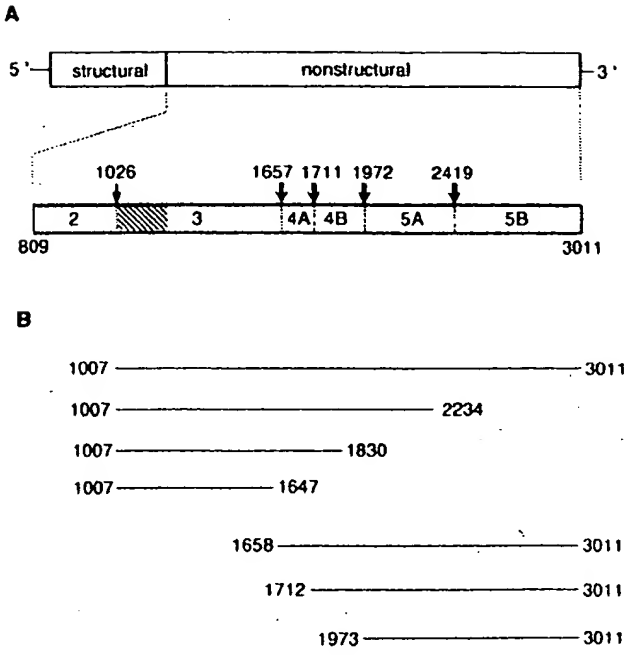


FIG. 1. HCV genome structure and expression constructs. (A) Diagram of the HCV genome encoding the structural proteins in the 5′-terminal quarter followed by the NS proteins 2 to 5B. The 5′ and 3′ nontranslated regions are indicated as thin lines. A detailed view of the NS protein region (amino acids 809 to 3011 of the polyprotein) is drawn below. Cleavage sites of the NS2-3 proteinase (↓) (16, 20) and the NS3 proteinase (⇓) (2, 13, 17, 20, 36) are given. Numbers above the arrows refer to the amino acids at the P1 positions of the scissile bonds. (B) HCV polyprotein expression constructs used in this study. Lines depict regions of the HCV genome expressed with recombinant vaccinia viruses and are drawn to scale and oriented with respect to the diagram in panel A. Numbers refer to the first and last amino acids of the HCV polyprotein expressed with the recombinant vaccinia viruses.

Metabolic labelling of infected cells and characterization of HCV-specific proteins. HeLa cells were infected at a multiplicity of infection of 5 to 10 as described above and 16 h later were incubated in methionine- and FCS-free medium for 1 h. After addition of the same medium supplemented with 100 μCi of [^{35}S]methionine (Amersham Life Science, Braunschweig, Germany) and incubation for various times, cells were lysed in TNE (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 1 mM EDTA) with 1% Triton X-100. The lysate was clarified by a 15-min centrifugation at 15,000 × *g* at 4°C, and proteins in the supernatant were precipitated by the addition of 2% sodium dodecyl sulfate (SDS) and 5% trichloroacetic acid. After 10 to 30 min at 0°C, precipitated proteins were collected by a 5-min centrifugation at 6,000 × *g* and resolved in protein sample buffer (200 mM Tris-HCl [pH 8.8], 5 mM EDTA, 2% SDS, 1%

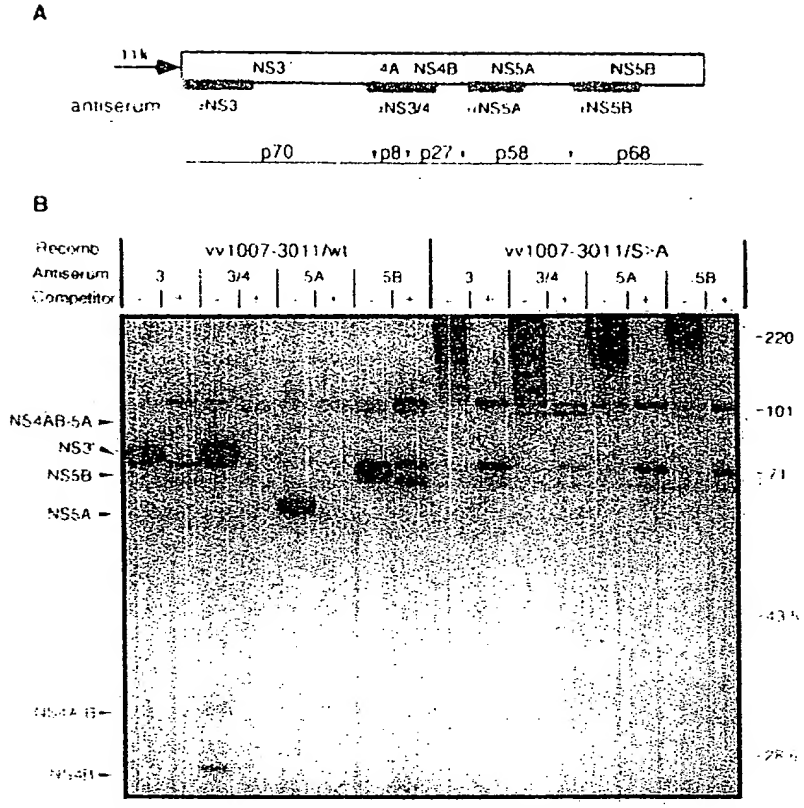


FIG. 2. Proteolytic processing of an NS3'-5B polyprotein expressed with recombinant vaccinia viruses. (A) Schematic representation of the HCV genome segment expressed under control of the vaccinia virus 11-kDa late promoter. Fragments of the polyprotein used to generate antisera are indicated as dotted bars. The apparent molecular masses (in kilodaltons) of individual processing products are given below. (B) Analysis of HCV-specific proteins isolated from HeLa cells infected with recombinants expressing an unaltered NS3'-5B polyprotein (vv1007-3011/wt) or an NS3'-5B with an enzymatically inactive NS3 proteinase (vv1007-3011/S→A). Cells were labelled for 1 h with [^{35}S]methionine, and proteins were isolated from the lysate by immunoprecipitation. To demonstrate the specificities of the detected proteins, immunoprecipitations were performed in the absence (−) or presence (+) of a homologous competitor. Numbers to the right refer to the sizes of marker proteins (in kilodaltons).

2-mercaptoethanol, 10% sucrose, and 0.1% bromophenol blue). After 5 min of boiling, samples were diluted to RIPA buffer by adding 20 volumes of RIPA buffer without SDS (PBS, 1% Triton X-100, 0.5% sodium deoxycholate) containing 1 mM phenylmethylsulfonyl fluoride (Sigma). Finally, 15 μl of packed protein A-Sepharose containing preadsorbed immunoglobulin (corresponding to 3 to 6 μl of antiserum) was added and the samples were incubated overnight at 4°C with agitation. After three washes of the immunocomplexes with RIPA buffer, protein sample buffer containing 3.3% SDS and 2% 2-mercaptoethanol was added, samples were boiled for 5 min, and half of the material was analyzed by SDS-polyacrylamide gel electrophoresis and fluorography. Competitions were done by adding 10 μg of purified homologous antigen to the immunoprecipitation.

In vitro transcription and translation. A detailed description of these methods is given in reference 2. In brief, the linearized plasmid pBSK 1007-1912 was used for in vitro transcription, and after phenol extraction and ethanol precipitation, RNA was quantified by comparison of a serial dilution of the RNA with a concentration standard after electrophoresis through an agarose gel. RNA was used for in vitro translations in a rabbit reticulocyte lysate according to the instructions of the manufacturer (Promega, Heidelberg, Germany).

RESULTS

Processing activity of an HCV NS3'-5B polyprotein. The goal of our studies was an examination of HCV polyprotein processing mediated by the NS3 serine-type proteinase by using an NS3'-5B polyprotein expressed with recombinant vaccinia viruses (Fig. 1) (a prime indicates an HCV protein with a nonauthentic amino or carboxy terminus). This polyprotein was selected because it was expressed to much higher levels than the full-length polyprotein (1a) and because it is predicted to properly reflect all the cleavage reactions mediated by the NS3 proteinase for two reasons: (i) we (2) and others (13, 17, 20, 36) have shown that sequences amino terminal to NS3 are dispensable for all NS3-mediated cleavages, and (ii) mutational ablation of the NS2-3 proteinase does not affect processing at any of the NS3-dependent cleavage sites (16, 20). The NS3'-5B polyprotein includes the complete NS3-4AB-5AB open reading frame and initiates with 20 amino acids derived from the carboxy terminus of NS2 (Fig. 2, vv1007-3011/wt). To correlate the generation of individual processing products with NS3-encoded activity, a further recombinant expressing an NS3'-5B polyprotein in which the putative active site serine residue was replaced by an alanine residue (vv1007-3011/S→A) was made (Fig. 2).

Infection of HeLa cells with recombinant virus expressing the active proteinase generated proteins with apparent molecular masses of about 70, 27, 58, and 68 kDa corresponding to NS3', NS4B, NS5A, and NS5B, respectively, thus demonstrating specific cleavage at the NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B sites (Fig. 2). Two additional double bands in the molecular mass range of about 67 and 95 kDa were consistently observed in all immunoprecipitations. They corresponded to vaccinia virus proteins binding nonspecifically to protein A-Sepharose since their amounts could not be reduced in immunoprecipitations using homologous competitors (Fig. 2, + lanes). We did not detect NS4A, a protein of about 8 kDa, possibly because of low reactivity with our antiserum and low labelling efficiency (the predicted NS4A of our isolate would contain only one methionine). However, the appearance of NS3 and NS4B can be taken as evidence of cleavage between NS3 and NS4A as well as between NS4A and NS4B.

In addition, various amounts of processing intermediates were detected and identified by their apparent molecular weights and their reactivities with the individual antisera as NS4AB and, much more visible in the experiments described below (e.g., see Fig. 3A), NS4AB-5A and NS4B-5A . None of these cleavage products was obtained with the polyprotein in which the active site serine residue was replaced by an alanine residue. Instead, the unprocessed polyprotein with an apparent molecular mass of about 220 kDa was detected, illustrating that the NS3-encoded proteinase is essential for cleavage at all sites downstream of its own carboxy terminus.

Kinetics of NS3'-5B polyprotein processing. To identify possible precursors, pulse-labelling experiments were carried

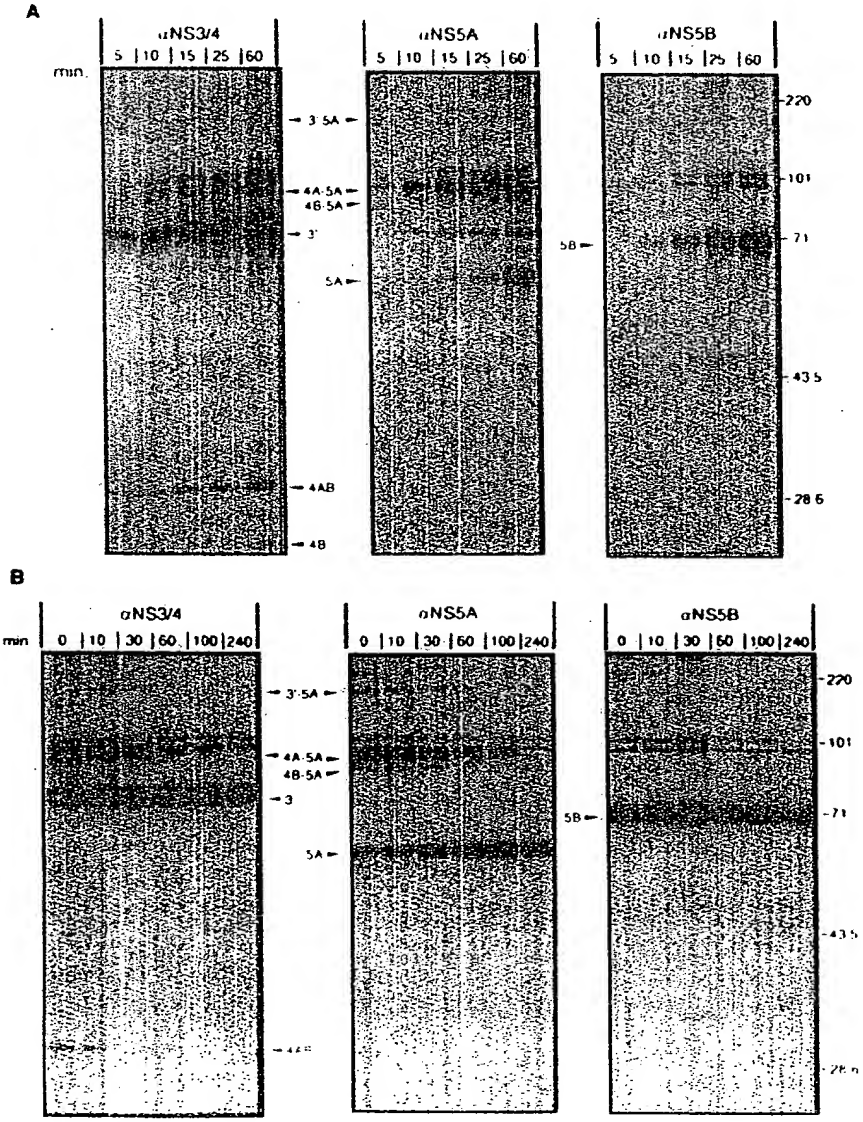


FIG. 3. Kinetic analyses of NS3'-5B polyprotein processing by continuous and pulse-chase labelling. (A) Cells were infected with the recombinant expressing NS3'-5B and 16 h later radiolabelled with [^{35}S]methionine for the indicated times. (B) Infected cells were labelled for 20 min and, after being washed, incubated in nonradioactive medium for various times. HCV-specific proteins were isolated from the cell lysate by immunoprecipitation with the indicated antisera.

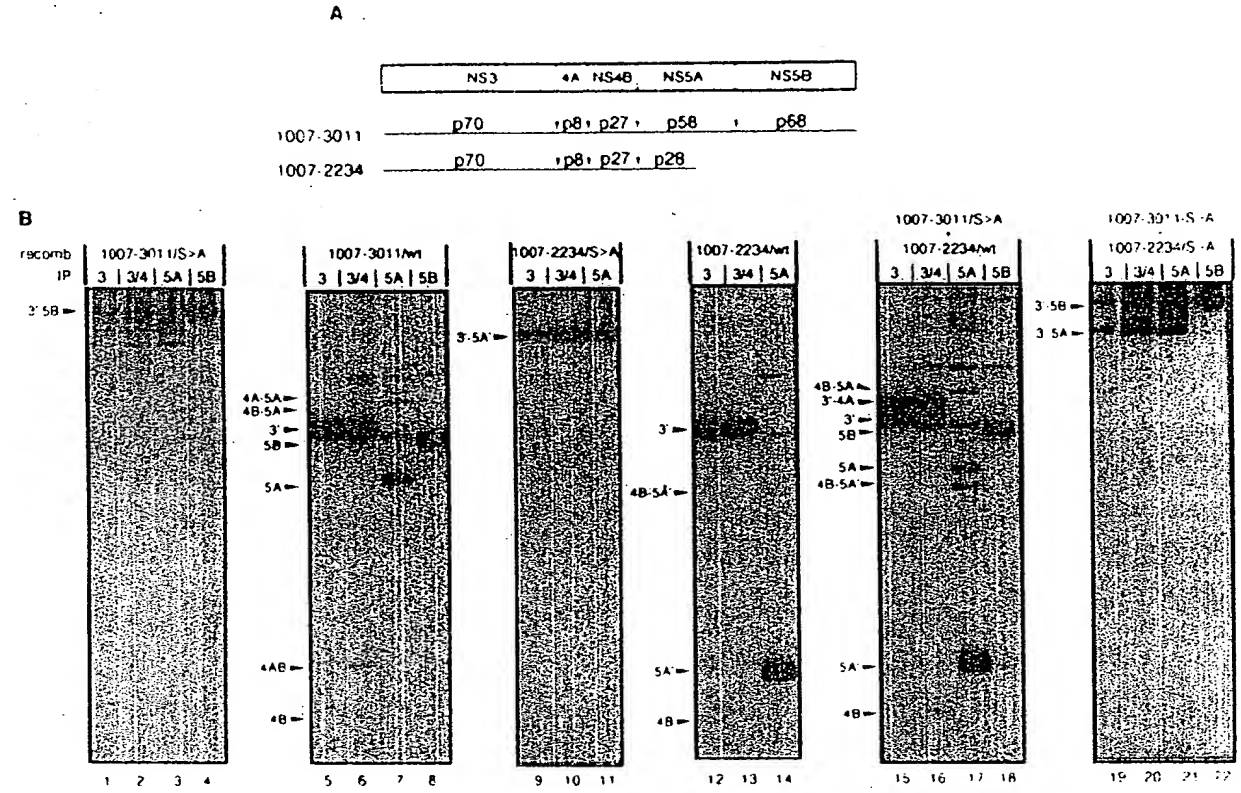


FIG. 4. Cleavage of an NS3'-5B polyprotein in *trans*. (A) Schematic representation of the HCV NS3'-5B and NS3'-5A' polyproteins and the apparent molecular weights of processing products. (B) Cells were infected with either one or two recombinants (recomb.) expressing unaltered HCV polyproteins (wt) or polyproteins with an inactive NS3 proteinase (S→A). Sixteen hours after infection, cells were radiolabelled for 1 h and then underwent cell lysis and immunoprecipitation (IP).

out with HeLa cells infected with vaccinia virus recombinants expressing the enzymatically active NS3'-5B polyprotein. The processed NS3' was already visible after 5 min of labelling (Fig. 3A, left panel); labelling times as short as 2 min still allowed clear detection of this protein (data not shown), suggesting that cleavage at the NS3/4A site is a rapid event. NS4B was detected after 60 min of continuous labelling (25 min with prolonged exposures), indicating delayed cleavage between NS4A and NS4B. In contrast, the NS4AB processing intermediate was generated very rapidly, being detectable after 5 min of labelling (Fig. 3A, left panel). NS5A and NS5B could be detected after 15 and 10 min of labelling, respectively. In addition to the mature processing products, several higher-molecular-mass proteins of about 85, 95, and 170 kDa were visible. Antiserum to NS3/4 and NS5A reacted with the 170 and 95 kDa proteins, whereas the 85 kDa protein was precipitated only by the NS5A-specific antiserum. None of these proteins reacted with the antiserum to NS5B. On the basis of these reactivities and deduced molecular masses, the 170-kDa protein can be designated NS3'-5A, the 95-kDa protein can be designated NS4AB-5A, and the 85-kDa protein can be designated NS4B-5A. No NS5B-specific processing intermediates were detected.

To analyze the stabilities of these proteins and to identify possible precursor-product relationships, pulse-chase experiments were performed (Fig. 3B). Results from immunoprecipitations with the NS3/4-specific antiserum revealed that NS3' forms a stable peptide with a half-life of about 4 h. In contrast, the NS3'-5A and NS4AB intermediates were unstable and disappeared with half-lives of about 45 and 15 min, respectively. The only protein with a clear precursor-product relationship was NS5A, whose level increased over time concomitantly with a decrease of NS4AB-5A and NS4B-5A intermediates. Using longer pulse intervals, we could also detect NS4B but found it to be very unstable, with a half-life of only about 15 min (data not shown). The only protein detected with the NS5B-specific antiserum was NS5B (Fig. 3B), a stable protein decaying with a half-life of about 6 h.

Processing of an NS3'-5B polyprotein in *trans*. So far we had used a polyprotein in which the proteinase was part of the polyprotein and therefore directly linked with its substrate. To differentiate whether processing at the various sites was strictly an intramolecular reaction or could also occur intermolecularly, transcomplementation experiments were performed (Fig. 4). Cells were infected with a recombinant vaccinia virus expressing the proteolytically inactive NS3'-5B polyprotein along with a recombinant directing the expression of an NS3'-5A' protein with an enzymatically active proteinase domain (Fig. 4A). As shown in Fig. 4B (lanes 15 to 18), this double infection yielded proteins corresponding to NS5A (lane 17) and NS5B (lane 18), both of which were not detected in cells expressing either inactive NS3'-5A' or NS3'-5B protein (lanes 9 to 11 and 1 to 4, respectively) nor in cells expressing both defective polyproteins together (lanes 19 to 22). A novel processing product of 78 kDa which was not produced by the active NS3'-5B polyprotein (Fig. 4B, lanes 5 to 8) was observed

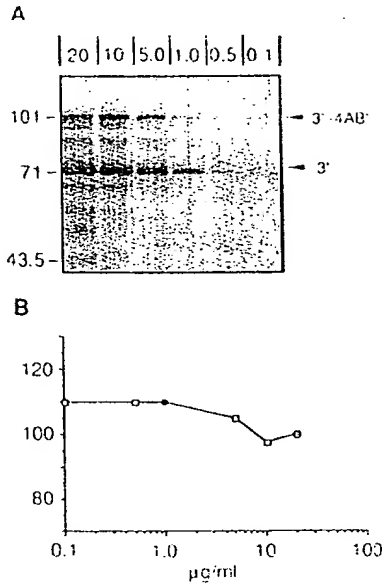


FIG. 5. Cleavage at the NS3/4A site is insensitive to dilution. (A) Various amounts of an HCV RNA encoding an NS3'-4B' polyprotein (amino acids 1007 to 1912) were used for in vitro translation with the rabbit reticulocyte lysate. After 60 min at 30°C, proteins were analyzed by electrophoresis in an 11% polyacrylamide gel and by fluorography. Numbers above each lane refer to the amount of RNA (in micrograms per milliliter) used for each translation. The positions of marker proteins (in kilodaltons) are indicated on the left. (B) Radioactivities contained in the NS3'-4B' and NS3' proteins were quantitated with a phosphoimager. The precursor product ratio obtained with 20 μg/ml was set as 100.

reacting with the NS3/4-specific antiserum (compare lanes 6 and 16). The size and immunoreactivity of this protein, together with the fact that no NS4AB-5A was observed with this infection but only an NS4B-5A intermediate (Fig. 4B; compare lanes 7 and 17), suggested that this polypeptide corresponded to an unprocessed NS3'-4A fusion protein. It should be noted that the labelling procedure required to monitor efficient *trans* cleavage and the kinetics of processing in *trans* (see below) did not allow clear detection of the unstable NS4B (half-life of about 15 min). However, the appearance of NS3'-4A and NS5A is taken as evidence of processing at the NS4A/4B and NS4B/5A junctions. Taken together, we found that cleavage between NS4A and NS4B, NS4B and NS5A, and NS5A and NS5B could be mediated in *trans*, whereas processing at the NS3/4A site appeared to occur essentially in *cis*, i.e., intramolecularly.

If cleavage between NS3 and NS4A is an intramolecular reaction, it should follow first-order reaction kinetics and be concentration independent. Therefore, we examined the protein products obtained by in vitro translation of an NS3'-4AB' RNA using different amounts of RNA to effectively change the concentration of the proteinase. As shown in Fig. 5A, translation of this RNA yielded unmistakable amounts of unprocessed precursor and the NS3' cleavage product, while NS4AB' could not be detected under these conditions (2). Quantitation of the radioactivities contained in these bands revealed no significant difference in the precursor-to-product ratio, irrespective of the RNA concentration (Fig. 5B). The same result was found with shorter and longer translation times (data not shown). These results together with the finding that cleavage between NS3 and NS4A appeared to occur in *cis* strongly suggested that processing at this site was an intramolecular reaction.

To analyze the kinetics of polyprotein processing in *trans*, pulse-chase experiments were performed with the proteolytically inactive NS3'-5B polyprotein as a substrate and with an NS3'-4B' proteinase (Fig. 6A). Cleavage was slower than processing of the enzymatically active NS3'-5B (Fig. 3) because obvious amounts of unprocessed substrate were detected even after a 2-h chase (Fig. 6B). Amounts of NS5B and NS5A increased with time, reaching maximum levels after about 1 and 2 h, respectively. Surprisingly, processing between NS4A and NS4B occurred with the fastest kinetics as shown by the rapid production of NS3'-4A and NS4B-5AB. This is in contrast to the results obtained with the enzymatically active NS3'-5B, when cleavage at this site was delayed (Fig. 3). It should be noted that essentially the same kinetics were found with the NS3'-5A' proteinase described above (data not shown), showing that sequences downstream of NS4B are dispensable for full proteolytic activity.

Much slower kinetics of *trans* cleavage and a slightly different pattern of processing products and processing intermediates were observed with an NS3 proteinase lacking NS4 sequences and 10 amino acids from the carboxy terminus of NS3 (vv1007-1647) (Fig. 6C). Compared with the results obtained with the NS3'-4B' proteinase, processing at the NS4A/4B site occurred with similar kinetics as shown by the rapid appearance of NS3'-4A and the NS4B-5AB intermediate. However, cleavage between NS5A and NS5B and, most notably, between NS4B and NS5A was much slower, and chase periods of 6 and 1 h were required to detect obvious amounts of NS5A and NS5B, respectively. Furthermore, an unmistakable accumulation of the NS4B-5A processing intermediate during the 6-h chase was observed. Since essentially the same results were found with an NS3 proteinase with the authentic carboxy terminus (amino acid 1657 of the polyprotein) (data not shown), sequences of the NS4 region coexpressed with the proteinase enhance processing efficiencies at all *trans*-cleavage sites, particularly between NS4B and NS5A (see below).

Mapping of the minimal NS3 domain required for proteolytic activity. Sequence comparisons between NS3 of HCV and other viral and nonviral proteinases suggest that the proteolytic activity is located in the amino-terminal domain of the molecule. Hence, to map the minimal NS3 domain required for polyprotein processing, a series of recombinant vaccinia viruses was generated, directing the expression of various NS3 proteins beginning at amino acid 1007 of the polyprotein and ending at different positions in the NS3 region. These recombinants were used for double infections along with the recombinant expressing the inactive NS3'-5B substrate polyprotein. Because of the slow reaction kinetics observed with proteinases lacking NS4 sequences, in this and all subsequent experiments, cells were radiolabelled for 1 h and then chased for 6 h. To exclude possible cleavage products generated by cellular or vaccinia virus-encoded enzymes, infections with wild-type vaccinia virus were included. As shown by the production of NS3'-4A, NS5A, and NS5B (Fig. 7A, B, and C, respectively), all variant NS3 proteinases could process the substrate at all *trans*-cleavage sites. In independent experiments, no significant differences with respect to processing efficiencies and kinetics were found between these NS3 truncations and the full-length NS3 proteinase (data not shown). Variations in the amounts of processing products were due to variations in the expression levels of the NS3 proteins and to their different stabilities. Most of them had half-lives much shorter than those observed

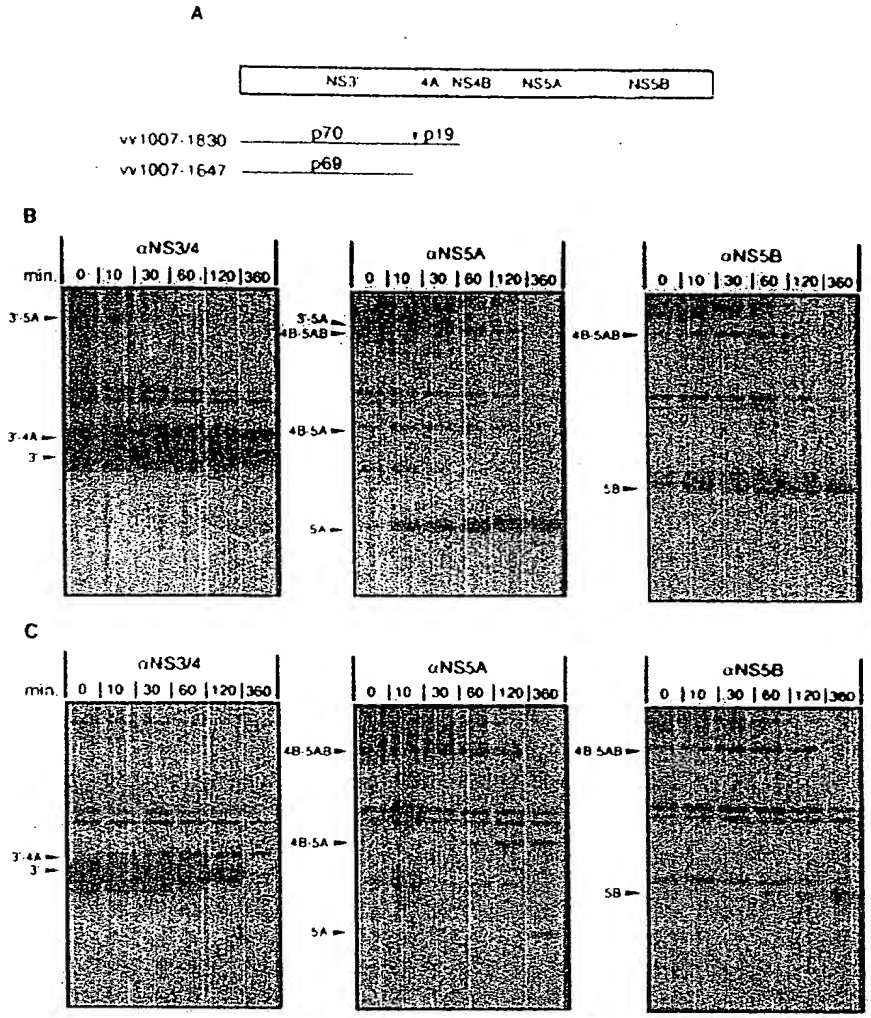


FIG. 6. Pulse-chase analysis of NS3'-5B polyprotein processing in *trans* by the NS3 proteinase in the presence and absence of NS4 sequences. (A) Schematic representation of the enzymatically inactive NS3'-5B polyprotein substrate and two variant proteinases. (B) Cells were infected with a combination of the recombinant expressing the inactive polyprotein and a recombinant expressing an enzymatically active NS3'-4B' proteinase (amino acids 1007 to 1830 of the polyprotein). Sixteen hours after infection, proteins were radiolabelled metabolically for 20 min and then incubated in nonradioactive medium for the indicated times. (C) Results from an analogous experiment with the same NS3'-5B substrate and an almost complete NS3 proteinase lacking NS4 sequences (amino acids 1007 to 1647 of the polyprotein).

for full-length and nearly full-length NS3 molecules and could be detected only in cell lysates prepared directly after the pulse-labelling (Fig. 7D). In summary, these results demonstrate that the first 211 amino acids of NS3 (expressed from recombinant vv1007-1238) are sufficient for processing at all *trans*-cleavage sites.

To determine the borders of the minimal NS3 proteinase domain more precisely, the smallest protein was used to introduce a series of amino-terminal deletions which were tested in the same way. Removal of only 7 amino acids from the amino terminus of the NS3 protein abolished cleavage at the NS4B/5A site (vv1034-1238) (Fig. 8B) without affecting processing at the other sites (Fig. 8A and C). A similar pattern was found when 23 amino-terminal amino acids were deleted (vv1050-1238). Deletion of an additional 16 amino acids from the NS3 amino terminus also abolished cleavage at the NS5A/5B site, whereas processing between NS4A and NS4B still occurred, albeit with lower efficiency (vv1066-1238). None of the cleavage products was obtained with an inactive NS3 proteinase (vv1050-1238/S→A), excluding the possibility that vaccinia virus or cellular enzymes could substitute for an active NS3 proteinase. These results show that deletions in the NS3 domain have differential effects on processing at the various cleavage sites and that cleavage at the NS4B/5A site is the most sensitive to amino-terminal deletions in the proteinase domain, whereas *trans* cleavage between NS4A and NS4B can be mediated by a proteinase as short as 172 amino acids from the amino-terminal region of NS3.

Further attempts to narrow down the minimal proteinase domain were complicated by the fact that these proteins probably did not react or reacted very poorly with our NS3-specific antiserum; therefore, the expression of these proteins could not be determined.

NS4 sequences are essential for cleavage at the NS4B/5A site. In all the *trans*-cleavage assays described so far, we used the NS3'-5B polyprotein as a substrate. To analyze whether

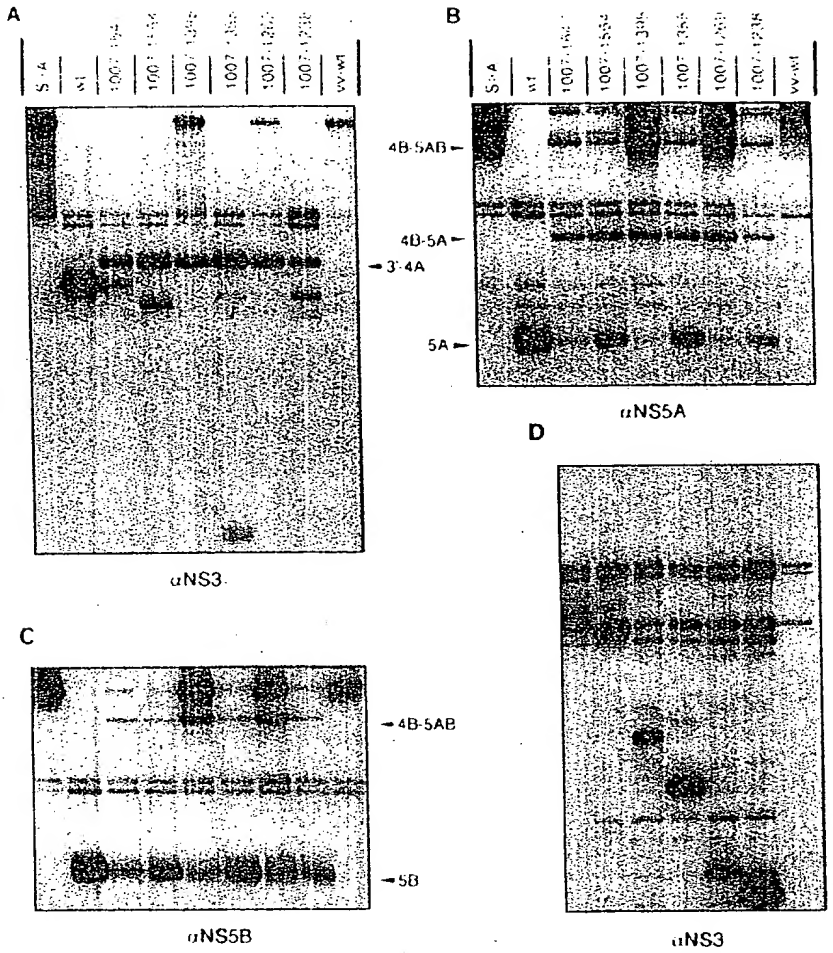


FIG. 7. Mapping of the minimal NS3 domain required for full proteolytic activity. Cells were infected either singly with one of the NS3 recombinants or with vaccinia wild-type virus (vv-wt) (D) or in combination with the recombinant expressing the inactive NS3'-5B substrate (A to C). Sixteen hours later, cells were labelled for 1 h and lysed directly (D) or after a 6-h chase (A to C). HCV-specific proteins were analyzed by immunoprecipitation with antisera specific for NS3 (A and D), NS5A (B), or NS5B (C). As a control, the results obtained with cells expressing an unaltered (wt) or mutated (S→A) NS3'-5B polyprotein are shown in the two lanes on the far left. Descriptions of the individual lanes in panels A and B also refer to the lanes directly below in panels C and D, respectively.

sequences of the substrate have an influence on cleavage at the various sites, recombinant vaccinia viruses expressing NS4A-5B, NS4B-5B, and NS5A-5B polyproteins were made, and we determined their processing patterns as generated by an NS3'-4B' proteinase (Fig. 9, vv1007-1830, II) or the NS3' proteinase lacking NS4 sequences and 10 amino acids from the NS3 carboxy terminus (Fig. 9, vv1007-1647, I).

trans cleavage of the NS4A-5B substrate with either of the proteinases gave rise to mature NS5A and NS5B, demonstrating cleavage at the NS4B/5A and NS5A/5B sites (Fig. 9A). Furthermore, small amounts of NS4B indicative of processing between NS4A and NS4B were detected (data not shown). Cleavage of the NS5A-5B polyprotein substrate by either of the proteinases occurred with similar efficiency (Fig. 9C). However, complete processing of the NS4B-5B substrate was observed only with the NS3'-4B' proteinase (II) (Fig. 9B), whereas in the case of the pure NS3 proteinase, cleavage occurred only at the NS5A/5B junction and no processing between NS4B and NS5A was detected (I) (Fig. 9B). The same result was found when we increased the amount of the NS3' proteinase by raising the multiplicity of infection of the vv1007-1647 recombinant. These results strongly suggest that NS4, particularly NS4A, is essential for processing between NS4B and NS5A and that NS4A provided by either the substrate (i.e., in *cis*) or the proteinase (i.e., in *trans*) can restore cleavage at this site.

DISCUSSION

The present study examined processing of the HCV NS proteins with the help of recombinant vaccinia viruses expressing an NS3'-5B polyprotein. This protein was chosen because it contains all the sequences required for NS3-mediated polyprotein processing and has considerably higher expression levels than the complete polyprotein (1a).

Results from kinetic studies revealed a complex pattern of NS polyprotein processing. The kinetics of processing between NS3 and NS4A were compatible with the view that cleavage at this site occurs first. However, in addition, small amounts of an NS3'-5A intermediate were detected, suggesting that alterna-

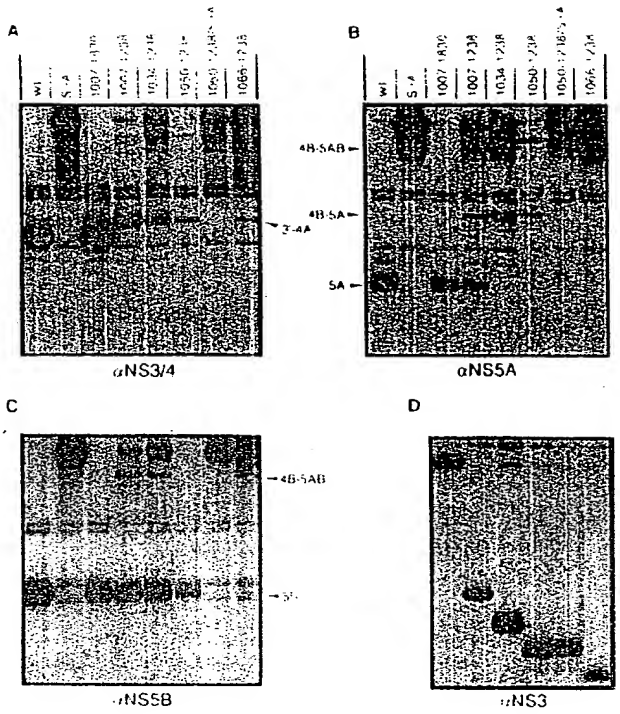


FIG. 8. Fine mapping of the minimal NS3 proteinase domain. Cells were doubly infected with the recombinant expressing the inactive NS3'-5B polyprotein substrate and one of several recombinants expressing various NS3 truncations. Following radiolabelling for 1 h, cells were lysed either directly (D) or after a 6-h incubation in nonradioactive medium (A to C). HCV-specific proteins were isolated by immunoprecipitation with the given antisera. Results obtained with cells expressing an unaltered (wt) or mutated (S→A) NS3'-5B polyprotein are shown in the two lanes on the far left. (D) To determine the expression of various NS3 truncations, cells were infected with the corresponding recombinants, labelled for 1 h, and analyzed by immunoprecipitation with the NS3-specific antiserum. Descriptions of the individual lanes in panels A and B also refer to the lanes directly below in panels C and D, respectively.

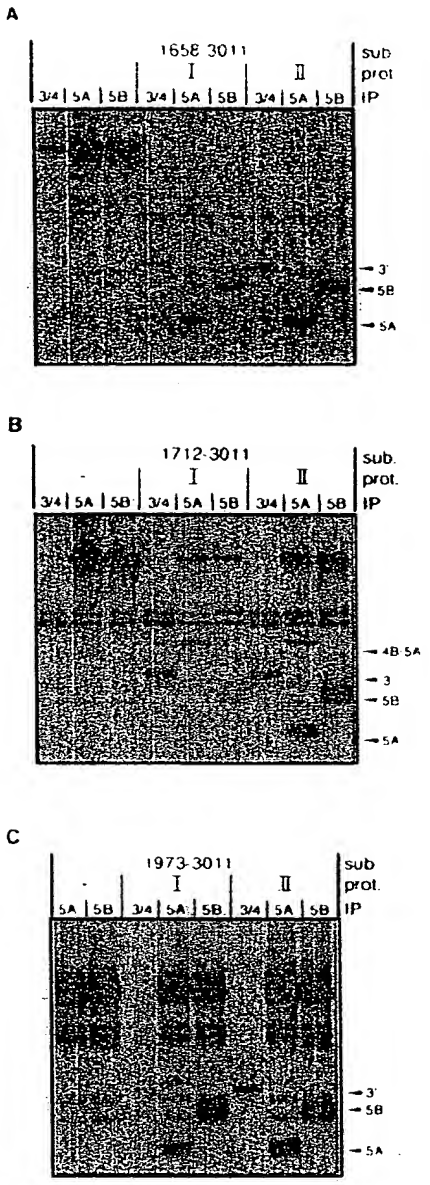


FIG. 9. Processing of NS4A-5B (A), NS4B-5B (B), and NS5A-5B (C) polyproteins (amino acids 1658 to 3011, 1712 to 3011, and 1973 to 3011 of the HCV polyprotein, respectively) in *trans* with recombinant vaccinia viruses expressing an almost complete NS3 proteinase (vv1007-1647) (I) or an NS3'-4B' proteinase (vv1007-1830) (II). For details, see the legend to Fig. 7.

tive cleavages such as initial processing at the NS5A/5B site exist. Cleavage between NS4A and NS4B appeared to be delayed as shown by the presence of the NS4A/B intermediate and the slow production of NS4B, which was detected only after 60 min of labelling compared with 2 min in case of NS3.

At least two cleavage pathways seem to operate at the NS4B/5A site: (i) rapid cleavage as indicated by rapid production of the NS4AB intermediate and (ii) slow cleavage as indicated by the presence of the relatively stable NS4AB-5A intermediate (half-life of ca. 1 h), for which a clear precursor-product relationship was found. In contrast, processing between NS5A and NS5B appears to be rather efficient because no precursor of NS5B could be detected. Thus, NS5B seems to be generated either cotranslationally from an unprocessed polyprotein by an intramolecular reaction or from an unstable processing intermediate which is cleaved very rapidly intermolecularly.

Different kinetics were observed in the *trans*-cleavage reaction. Processing at the NS3/4A site could not be detected, and cleavage was fastest at the NS4A/4B site. A possible explanation for this apparent difference is that in the case of the enzymatically active NS3'-5B polyprotein, a transient intramolecular complex between the NS3 proteinase domain and NS4 sequences, particularly NS4A, forms, which brings the cleavage site in close proximity to the active site of the enzyme to ensure rapid processing between NS3 and NS4A. Consistent with this idea, we obtained strong evidence that cleavage at this site occurs in *cis*. However, in the case of the inactive NS3'-5B polyprotein, the proteinase domain would remain stably associated with NS4, making the cleavage site inaccessible to proteinase molecules provided in *trans*. Hence, one might expect processing at this site to occur in *trans* with a polyprotein lacking the NS3 proteinase domain. We are currently testing this hypothesis.

Using a genetic approach, we found that sequences from the NS4 region, particularly NS4A, are important for efficient polyprotein processing, especially at the NS4B/5A site. Cleavage at this site was accomplished when NS4A was present in the substrate (in *cis*) and even more efficiently when it was coexpressed with the proteinase (in *trans*). The mechanism by which NS4A promotes cleavage is currently not known. Several possibilities can be envisaged. (i) NS4A might be important for proper folding of the substrate. This possibility seems unlikely because NS4 can be provided in *trans* to restore cleavage and the NS5 substrate lacking NS4 sequences is processed properly. (ii) NS4A may form a complex with NS3 that is important for NS3 function and analogous to the NS2B/3 heterodimer described for flaviviruses (1). However, in contrast to those viruses, in the case of HCV, formation of such a complex would not be a prerequisite for proteinase activity per se because an NS3 proteinase lacking NS4 sequences still could cleave at the NS5A/5B junction. (iii) NS4A may serve as a membrane anchor attaching the more hydrophilic NS3 proteinase to the membrane surface of the endoplasmic reticulum, where most of the HCV proteins appear to be located (21, 30). In this model, NS4A would bring the NS3 proteinase in close proximity to its substrate, allowing efficient processing, particularly at the NS4B/5A junction. Sequence analysis of NS4A reveals that the first half of the molecule is very hydrophobic and has the potential to span the lipid bilayer once, whereas the other half of the molecule is very hydrophilic (5 of 10 carboxy-terminal amino acids are acidic) and may extend into the cytoplasm to interact with the more hydrophilic NS3. In agreement with this idea, Hijikata and coworkers (21) have recently shown that membrane association of NS3 translated in vitro in the presence of microsomal membranes requires the presence of NS4A. It should be noted that this kind of interaction may contribute not only to efficient polyprotein processing but also to the formation of a putative membrane-associated replication complex (postulated by analogy to flaviviruses [7]) involving the nucleoside triphosphatase and helicase activities residing in the carboxy-terminal domain of NS3 (34).

In an attempt to narrow down the minimal NS3 domain required for full proteolytic activity, we found that the first 211 amino acids of NS3 suffice for cleavage at all *trans* sites. Further amino-terminal deletions selectively abolished processing between NS4B and NS5A while cleavage at the other *trans* sites was not affected. Interestingly, the same phenotype, the loss of cleavage ability at the NS4B/5A site, was caused by the lack of NS4A protein. On the basis of these results, one might postulate specific interactions between the NS3 amino terminus and the NS4A protein. However, the existence and nature of such interactions remain to be determined.

For many viruses, proteolytic processing mediated by virus-encoded proteinases plays an important role for virus replication. Recently, it was shown that mutational ablation of the NS3-encoded proteinase of yellow fever virus severely reduced virus replication (8). Thus, it seems likely that the analogous proteinase of HCV plays an equally important role for the viral life cycle. Although we are still far away from a complete understanding of the pathways governing HCV polyprotein processing, our results are a first step in this direction and may provide a basis for the rational design of an antiviral drug.

ACKNOWLEDGMENTS

We thank K. Yasargil for excellent technical assistance, H. Jacot and W. Suppiger for immunization and care of rabbits, A. Wyss for synthesis of oligonucleotides, M. Hänggi and R. Wittek for the gift of vaccinia virus wild-type DNA, and D. Stüber for the gift of the pDS vector.

REFERENCES

1. **Arias, C. F., F. Preugschat, and J. Strauss.** 1993. Dengue virus 2 NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. Virology 193:888–899.
1a. **Bartenschlager, R.** Unpublished data.
2. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1993. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J. Virol. 67:3835–3844.
3. **Bartenschlager, R., C. Kuhn, and H. Schaller.** 1992. Expression of the P-protein of the human hepatitis B virus in a vaccinia virus system and detection of the nucleocapsid-associated P-protein by radiolabelling at newly introduced phosphorylation sites. Nucleic Acids Res. 2:195–202.
4. **Bukh, J., R. H. Purcell, and R. H. Miller.** 1993. At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc. Natl. Acad. Sci. USA 90:8234–8238.
5. **Cahour, A., B. Falgout, and C.-J. Lai.** 1992. Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. J. Virol. 66:1535–1542.
6. **Chambers, T. J., A. Grakoui, and C. M. Rice.** 1991. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J. Virol. 65:6042–6050.
7. **Chambers, T. J., C. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44:649–688.
8. **Chambers, T. J., R. C. Weir, A. Grakoui, D. McCourt, J. F. Bazan, R. J. Fletterick, and C. M. Rice.** 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA 87:8898–8902.
9. **Chien, D., Q. Choo, A. Tabrizi, C. Kuo, J. McFarland, K. Berger, C. Lee, J. Shuster, T. Nguyen, D. Moyer, M. Tong, S. Furuta, M. Omata, G. Tegtmeier, H. Alter, E. Schiff, L. Jeffers, M. Houghton, and G. Kuo.** 1992. Diagnosis of hepatitis C virus (HCV) infection using an immunodominant chimeric polyprotein to capture circulating antibodies: reevaluation of the role of HCV in liver disease. Proc. Natl. Acad. Sci. USA 89:10011–10015.
10. **Choo, Q.-L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradly, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B hepatitis genome. Science 244:359–362.
11. **Collett, M.** 1992. Molecular genetics of pestiviruses. Comp. Immunol. Microbiol. Infect. Dis. 15:145–154.
12. **Colombo, M., G. Kuo, Q.-L. Choo, M. F. Donato, E. D. Ninno, M. A. Tommasini, N. Dioguardi, and M. Houghton.** 1989. Prevalence of antibodies to hepatitis C virus in Italian patients with hepatocellular carcinoma. Lancet ii:1006–1008.
13. **Eckart, M. R., M. Selby, F. Masiarz, C. Lee, K. Berger, K. Crawford, C. Kuo, G. Kuo, M. Houghton, and Q.-L. Choo.** 1993. The hepatitis C virus encodes a serine proteinase involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. Biochem. Biophys. Res. Commun. 192:399–406.
14. **Falgout, B., M. Pethel, Y.-M. Zhang, and C.-J. Lai.** 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65:2467–2475.
15. **Gerber, M. A.** 1993. Relation of hepatitis C virus to hepatocellular carcinoma. J. Hepatol. 17(Suppl. 3):108–111.
16. **Grakoui, A., D. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** 1993. A second hepatitis C virus-encoded proteinase. Proc. Natl. Acad. Sci. USA 90:10583–10587.
17. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** 1993. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-depen-

dent polyprotein cleavage sites. J. Virol. 67:2832–2843.
18. **Harada, S., Y. Watanabe, K. Takeuchi, T. Suzuki, T. Katayama, Y. Takebe, I. Saito, and T. Miyamura.** 1991. Expression of processed core protein of hepatitis C virus in mammalian cells. J. Virol. 65:3015–3021.
19. **Hijikata, M., N. Kato, Y. Ootsuyama, and M. Nakagawa.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc. Natl. Acad. Sci. USA 88:5547–5551.
20. **Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno.** 1993. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J. Virol. 67:4665–4675.
21. **Hijikata, M., H. Mizushima, Y. Tanji, Y. Komoda, Y. Hirowatari, T. Akagi, N. Kato, K. Kimuran, and K. Shimotohno.** 1993. Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. Proc. Natl. Acad. Sci. USA 90:10773–10777.
22. **Hollinger, F. B.** 1990. Non-A, non-B hepatitis viruses, p. 2239–2273. *In* B. N. Fields (ed.), Virology. Raven Press, New York.
23. **Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo.** 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology 14:381–388.
24. **Hsu, H. H., M. Donets, H. Greenberg, and S. M. Feinstone.** 1993. Characterization of hepatitis C virus structural proteins with a recombinant baculovirus expression system. Hepatology 17:763–771.
25. **Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno.** 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA 87:9524–9528.
26. **Matsuura, Y., S. Harada, R. Suzuki, Y. Watanabe, Y. Inoue, I. Saito, and T. Miyamura.** 1992. Expression of processed envelope protein of hepatitis C virus in mammalian and insect cells. J. Virol. 66:1425–1431.
27. **Miller, R. H., and R. H. Purcell.** 1990. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus subgroups. Proc. Natl. Acad. Sci. USA 87:2057–2061.
28. **Preugschat, F., C.-W. Yao, and J. H. Strauss.** 1990. In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J. Virol. 64:4364–4374.
29. **Saito, I., T. Miyamura, A. Ohbayashi, H. Harada, T. Katayama, S. Kikuchi, Y. Watanabe, S. Koi, M. Onji, Y. Ohta, Q.-L. Choo, M. Houghton, and G. Kuo.** 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA 87:6547–6549.
30. **Selby, M. J., Q.-L. Choo, K. Berger, G. Kuo, E. Glazer, M. Eckart, C. Lee, D. Chien, C. Kuo, and M. Houghton.** 1993. Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. J. Gen. Virol. 74:1103–1113.
31. **Simmonds, P., E. C. Holmes, T.-A. Cha, S.-W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Ureda.** 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J. Gen. Virol. 74:2391–2399.
32. **Spaete, R. R., D. Alexander, M. Rugroden, Q.-L. Choo, K. Berger, K. Crawford, C. Kuo, S. Leng, C. Lee, R. Ralston, K. Thudium, J. W. Tung, G. Kuo, and M. Houghton.** 1992. Characterization of the hepatitis C virus E2/NS1 gene product expressed in mammalian cells. Virology 188:819–830.
33. **Stueber, D., H. Matile, and G. Garotta.** 1990. System for high-level production in Escherichia coli and rapid purification of recombinant proteins: application to epitope mapping, preparation of antibodies, and structure-function analysis, p. 121–152. *In* I. Lefkovits and B. Pernis (ed.), Immunological methods, vol. IV. Academic Press, New York.
34. **Suzich, J. A., J. K. Tamura, F. Palmer-Hill, P. Warrener, A. Grakoui, C. M. Rice, S. M. Feinstone, and M. S. Collett.** 1993. Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. J. Virol. 67:6152–6158.
35. **Tanaka, K., T. Hirohata, S. Koga, K. Sugimachi, T. Kanematsu, F. Ohryohji, H. Nawata, H. Ishibashi, Y. Maede, H. Kiyokawa, K. Tokunaga, Y. Irita, S. Takeshita, Y. Arase, and N. Nishino.** 1991. Hepatitis C and hepatitis B in the etiology of hepatocellular carcinoma in the Japanese population. Cancer Res. 51:2842–2847.
36. **Tomei, L., C. Failla, E. Santolini, R. de Francesco, and N. La Monica.** 1993. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. J. Virol. 67:4017–4026.
37. **Wiskerchen, M., and M. S. Collett.** 1991. Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. Virology 184:341–350.

Protein Expression and Purification 16, 440–447 (1999)
Article ID prep.1999.1090, available online at http://www.idealibrary.com on IDEAL®

An Uniquely Purified HCV NS3 Protease and NS4A$_{21-34}$ Peptide Form a Highly Active Serine Protease Complex in Peptide Hydrolysis

Vinod. V. Sardana,[1] Jeffrey T. Blue, Joan Zugay-Murphy, Mohinder K. Sardana,* and Lawrence C. Kuo

*Department of Antiviral Research and *Department of Biological Chemistry, Merck Research Laboratories, West Point, Pennsylvania 19486-0004*

Received March 12, 1999, and in revised form April 22, 1999

The N-terminal domain of the hepatitis C virus (HCV) polyprotein containing the NS3 protease (residues 1027 to 1206) was expressed in *Escherichia coli* as a soluble protein under the control of the T7 promoter. The enzyme has been purified to homogeneity th cation exchange (SP-Sepharose HR) and heparin affinity chromatography *in the absence of any detergent*. The purified enzyme preparation was soluble and remained stable in solution for several weeks at 4°C. The proteolytic activity of the purified enzyme was examined, also in the absence of detergents, using a peptide mimicking the NS4A/4B cleavage site of the HCV polyprotein. Hydrolysis of this substrate at the expected Cys–Ala scissile bond was catalyzed by the recombinant protease with a pseudo second-order rate constant (k_{cat}/K_M) of 205 and 196,000 M^{-1} s^{-1}, respectively, in the absence and presence of a central hydrophobic region (sequence represented by residues 21 to 34) of the NS4A protein. The rate constant in the presence of NS4A peptide cofactor was two orders of magnitude greater than reported previously for the NS3 protease domain. A significantly higher activity of the NS3 protease–NS4A cofactor complex was also observed with a substrate mimicking the NS4B/5A site (k_{cat}/K_M of 5180 ± 670 M^{-1} s^{-1}). Finally, the optimal formation of a complex between the NS3 protease domain and the cofactor NS4A was critical for the high proteolytic activity observed. © 1999 Academic Press

Human hepatitis C virus (HCV)[2] is the major etiologic agent of post transfusion non-A, non-B hepatitis (1,2). Chronic infection with HCV has also been linked to the development of liver cirrhosis and of hepatocellular carcinoma (3). Thus far, no efficient therapy exists and there is an urgent need for the development of HCV-specific antiviral therapeutics.

Along with flaviviruses and pestiviruses, HCV is a member of the *flaviviridae* family. These viruses share similarities in their genomic and polyprotein organization. HCV contains a positive-sense linear RNA genome of 9.5 kb with a single open reading frame encoding a polyprotein of 3010 to 3033 amino acids (4–7). This polyprotein encodes at least 9 different proteins as follows: 5′-C-E1-E2-NS2-NS3-NS4A-NS4B-NS5A-NS5B-3′ (where E denotes envelope proteins and NS denotes nonstructural proteins) and is proteolytically processed in the cytoplasm and/or in the endoplasmic reticulum. Both viral- and host-encoded proteases are involved in the production of mature viral proteins (8–10): First, the host signal peptidase appears to be responsible for cleavages in the structural NS2 region, generating C, E1, E2, and possibly NS2; second, the NS2/3 junction is cleaved by a putative HCV-encoded metalloprotease residing in the NS2 region; and finally, a second viral protease, essential for processing most of the NS proteins, is located in NS3. It has been shown that the N-terminal domain of NS3 encodes a serine protease that is necessary but not sufficient for efficient cleavages downstream of NS3, i.e., at the NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B junctions. The NS4A is an amphipathic protein of 54 amino acids, and

[1] To whom correspondence and reprint requests should be addressed. Fax: 215-983-3943. E-mail: vinod_sardana@merck.com.

[2] Abbreviations used: HCV, hepatitis C virus; DTT, dithiothreitol; DMSO, dimethyl sulfoxide; IPTG, isopropyl-D-(−)-thiogalactopyranoside; Nph, *para*-nitrophenylalanine; NS, nonstructural.

1046-5928/99 $30.00

it has a hydrophobic N-terminal domain followed by a hydrophilic C-terminal domain. The NS4A acts as a cofactor of the NS3 protease activity for an efficient cleavage of NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B sites and it interacts with the NS3 protease *in trans*. Recently, it has been suggested that the central region encompassing residues 21–34 of NS4A mimics the NS4A in its activation of NS3 *in vitro* (11–19). However, little is known about the mechanism by which the intact NS4A regulates proteolytic activity of the NS3 protease.

The 20-kDa N-terminal domain of the NS3 protein is capable of catalyzing hydrolysis of the cleavage site downstream from NS3 with the same efficiency as the full-length NS3 protein; it also retains its ability to interact with NS4A (11). Recently, several groups have reported purification of full-length NS3 protease as a histidine fusion protein (20) and as a maltose binding protein (MBP) fusion protein (15,21) from *Escherichia coli*. The amino terminal domain of NS3 protease by itself or as a fusion protein has also been purified from *E. coli* (22–24) and baculovirus (25,26) in the presence of detergents. These enzyme preparations exhibited very low activity for all peptide substrates examined even in the presence of NS4A cofactor (15,22,25,27,28).

In this report, we describe a unique and rapid two-step procedure for the purification of large quantities of the protease domain as a 19-kDa recombinant protein of NS3 protease from *E. coli* in the complete absence of detergents. Purification, substrate cleavage assays, and screening of compounds as inhibitors for HCV protease are simpler without interference from detergents. We further show that complex formation between the protease and the NS4A cofactor is important under the assay conditions for the dramatic activation of the enzyme by the 4A cofactor peptide to exhibit cleavage efficiency almost 1000-fold greater than the uncomplexed form, and at least 130-fold greater than those reported previously with peptide substrates.

MATERIALS AND METHODS

Expression of the HCV NS3 protease. The DNA encoding amino acids 1027–1206 (21) of the BK strain HCV polypeptide was cloned downstream of the T7-7 vector, in frame with the first ATG of the protein of gene 10 of the T7 phage, to yield the plasmid pT7-7(NS3$_{1027-1206}$). This plasmid was transformed into *E. coli* BL21DE3 plysS cells (Novagen) utilizing heat shock techniques. Cells were grown at 37°C in LB medium containing 50 μg/ml ampicillin to an optical density of 0.4–0.6 at 600 nm, and the temperature was lowered to 25°C to allow for induction with 400 μM isopropyl-D-(−)-thiogalactopyranoside (IPTG; Boehringer Mannheim). The bacterial cells were grown further for two additional hours and then pelleted with centrifugation for storage at −80°C.

Purification of HCV NS3 protease. Cells from a 10-liter culture were resuspended at 4°C in 100 ml of lysis buffer (25 mM sodium phosphate, pH 7.5, 1 mM EDTA, 10% glycerol, 5 mM DTT) and treated for 30 min with 0.02 mg/ml DNase (Type IIS: bovine pancreas, Sigma) in 20 mM $MgCl_2$. PMSF (1 mM) was added to the cell suspension and the cells were immediately disrupted with its passage six times through a Microfluidizer (Model 110-S) at 6 bar pressure. The cell lysate was centrifuged at 10,000 rpm for 30 min, and the supernatant was loaded at 2.5 ml/min onto a Hi-Load SP Sepharose high-performance column (26/10; Pharmacia Biotech) which had been preequilibrated in 50 mM sodium phosphate (pH 6.5), 10% glycerol, 1 mM EDTA, 5 mM DTT. The enzyme was eluted from the column in a 0–1 M NaCl gradient. Fractions collected were analyzed with sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and those containing proteins with molecular mass similar to the NS3 protease were pooled and diluted 8- to 10-fold into a buffer containing 25 mM sodium phosphate (pH 7.5), 10% glycerol, 5 mM DTT buffer, and loaded at 3 ml/min onto four columns of 5 ml Hi-trap heparin (Pharmacia-Biotech) connected in series without additional tubing. The enzyme was then eluted with a 0–1 M NaCl gradient. Fractions were analyzed by SDS–PAGE and with the peptide cleavage assays. Protein fractions demonstrating the NS3 enzymatic activity and showing greater than 99% purity on the SDS–PAGE were pooled and stored at −80°C in the elution buffer. The N-terminal sequence analysis was carried out by Edman degradation on an Applied Biosystems Model 494A protein sequencer. Protein concentrations were determined with quantitative amino acid analyses by using a postcolumn ninhydrin derivatization method on a Beckman 6300 analyzer.

Substrate cleavage assay. The peptides 7-methoxycoumarin-4-acetyl-DEMEECASHLPYK-(ε-NHCOCH$_3$) and acetyl-DEMEECASHLPYK-(ε-NHCOCH$_3$), mimicking the NS4A/4B cleavage site of the HCV polyprotein, were custom synthesized for us by Enzyme Systems Products (Dublin, CA) and were >95% pure. The NS4B/5A substrate 7-methoxycoumarin-4-acetyl-EDASTPCSGS-Nph-L (where Nph is *para*-nitrophenylalanine) was purchased from Bachem Biosciences. The NS4A peptide with the sequence of GSVVIVGRIILSGRKK [4Apep(21–34)KK] was also synthesized by Enzyme System Products. Peptide hydrolytic assays catalyzed by the NS3 protease were performed at 25°C in a circulation water bath in a buffer containing 100 μl of 50 mM Hepes (pH 7.5) and 10 mM DTT with 50% glycerol. The reaction was quenched with an aliquot of

100 μl of 5% phosphoric acid, and the mixture was analyzed with reverse-phase high-performance liquid chromatography (HPLC) on a 4.6 × 50-mm Vydac C18 column. The cleavage products were separated using a 0.1% trifluoroacetic acid/acetonitrile gradient and identified with comparisons of retention times against defined peptide products. Absorbance of the eluent was monitored at 220 nm using a Waters 996 photodiode array detector or a Waters 470 fluorescence detector. Excitation wavelength was set at 328 nm and the fluorescence emission was monitored at 393 nm. The enzyme concentrations used in the assays varied from 2 to 20 nM in the presence of 4Apep(21–34)KK and 500–1000 nM in the absence of the 4Apep(21–34)KK peptide. In the assays in which the 4Apep(21–34)KK was present, the enzyme was preincubated with 4Apep(21–34)KK for 10 min at 4°C, followed by 5 min at room temperature at a 20- to 40-fold greater concentration, before addition to the assay reaction mixture. For preincubation of enzyme with the 4A peptide, the enzyme was added to the solution already containing the 4A peptide (due to instability of NS3 by itself at the preincubation concentrations). The NS4A peptide and substrate concentrations used ranged from 0.1 μM to 50 μM and 0.25 μM to 100 μM, respectively. All substrates were dissolved in 50 mM Hepes (pH 7.5), 30 mM DTT and 10% glycerol. The assay was typically conducted for a period of 5 min in the absence of 4A peptides, but for less than 5 min when 4Apep(21–34)KK was present. The steady-state kinetic parameters (k_{cat} and K_M) were determined by fitting the initial-velocity versus substrate-concentration data to the Michaelis–Menten equation. Initial velocity and steady-state conditions were strictly maintained for all reaction assays performed.

Stability of NS3 protease domain 4Apep(21–34)KK complex. The complex was formed by mixing the NS3 protease domain (20 μM stock solution) and 4Apep(21–34)KK (1 mM stock) to a final concentration of 100 nM and 25 μM, respectively, in 50 mM Hepes (pH7.5), 10 mM DTT, and 50% glycerol, and incubated at 25°C. Activity was monitored as a function of time over 24 h.

RESULTS AND DISCUSSION

Expression and purification. The HCV protease was found to be localized in inclusion bodies when the plasmid DNA encoding the HCV protease was transformed into BL21DE3 pLysS *E. coli* cells and induced with IPTG at 37°C,. However, lowering the induction temperature to 25°C immediately following addition of IPTG and allowing the cells to grow for only 2 h resulted in the accumulation of HCV protease in the soluble fractions prepared from the host cells. Treatment with DNase prior to cell disruption rendered the solution less viscous upon cell lysis but it had no effect on solubility of the recombinant NS3 protease. We used a rapid two-step purification scheme for the purification of the protease. The supernatant obtained subsequent to centrifugation of cellular particulate was fractionated by SP Sepharose HP and heparin affinity chromatography and the eluent was analyzed by SDS-PAGE. A significant purification (>60% purity) was obtained with the SP Sepharose HP column. The second step of the Hi-trap heparin column yielded further purity to homogeneity (>99%) as shown in Fig 1. The two-step procedure described here was rapid and resulted in a homogenous preperation of NS3 protease without using any detergent during the whole purification procedure.

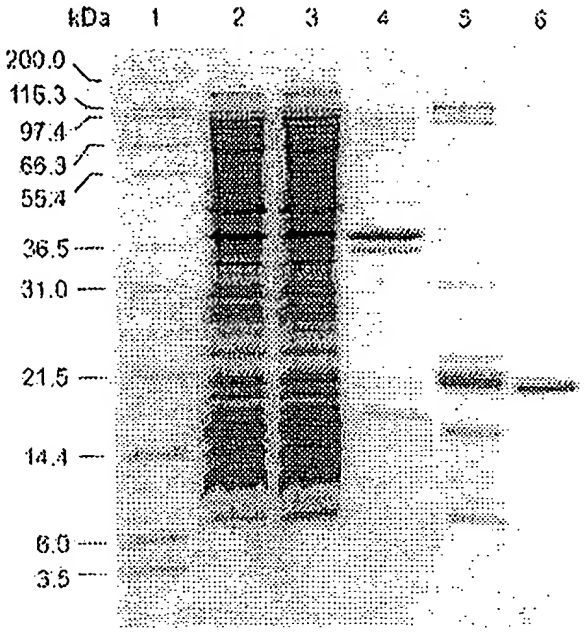


FIG. 1. Reduced SDS–PAGE (16%) analysis of the NS3 protease domain of human hepatitis C virus. The gel was stained with Coomassie brilliant blue. Lane 1, molecular weight markers. Lane 2, crude bacterial lysate. Lane 3, supernatant of crude lysate. Lane 4, pellet from crude lysate. Lane 5, pooled SP Sepharose HP fractions. Lane 6, pooled Hi-trap heparin fractions.

The purification of NS3 protease is summarized in Table 1. As judged with the enzymatic specific activity, a 3100- and 40,700-fold increase in activity was obtained with the SP Sepharose HP and Hi-trap-heparin columns, respectively. The initial activity observed may be low because the enzyme may be inhibited in the crude supernatant. The final yield of pure enzyme was 1 mg per liter of *E. coli* cell culture. Amino acid sequence analysis of the purified enzyme revealed that the N-terminal of the purified NS3 protease to be P-I-T-A-. . . instead of the expected M-A-P-I-T-A-. . . as deduced (6) from the cDNA sequence. Mass spectrometry data also confirmed this result—a molecular mass of

TABLE 1

Purification of HCV NS3 Protease Domain

Fraction	Protein (mg/ml)	Total protein (mg)	Specific activity (nmol · min^{-1} mg^{-1})	Fold purification
Crude supernatant	11.5	1380	0.001[a]	—
HR Sepharose	1.7	47.6	3.10	3100
Heparin	0.3	11.1	40.7	40,700

[a] This value is only a rough estimate of the upper limit of specific activity. For this reason, yield of purified NS3 protease is not included in this table.

18,867.8 (predicted average mass 18,868.6) was found for the purified enzyme. This value is in agreement with a NS3 protease domain containing amino acid residues from 1029 to 1206 without the encoded N-terminal Met and Ala residues as shown by N-terminal sequence analysis. Thus, the recombinant NS3 isolated from *E. coli* in these and other (22) studies had deletion of the first two N-terminal amino acids. The underlying mechanism of the cleavage is not clear.

In the past, several groups have reported purification of either the full-length or the N-terminal protease domain of NS3—from various sources such as *E. coli*, baculovirus, and mammalian host cells—and have emphasized (20,22–27,29) the need of adding various detergents to solubilize the protease in their purification procedures. In the current protocol, however, detergent is *not* necessary for the enzyme to be soluble throughout the course of the purification. Furthermore, the enzyme could be easily concentrated to 4–5 mg/ml and stored at 4°C for weeks without any apparent loss in activity.

Enzyme activity. To characterize the enzyme purified without detergent, first we developed a peptide cleavage assay using as substrate two peptides, both corresponding to the P6-P'6 residues of the NS4A/4B cleavage site. To one (Ac-DEMEECASHLPYK-(ε-NHCOCH$_3$); peptide I), a lysine was introduced to the C-terminus to render it more soluble at higher concentrations. To the other (7-methoxycoumarin-4-acetyl-DEMEECASHLPYK-(ε-NHCOCH$_3$); peptide II), a coumarin fluorophore was introduced at the N-terminus to enhance detection of the cleaved product.

The purified NS3 protease catalyzes the hydrolysis of both NS4A/4B substrates at the expected Cys-Ala scissile bond. In the absence of glycerol or at low glycerol concentrations, the rate of hydrolysis of the NS4A/4B substrate as catalyzed by the protease was very low but could be detected at high concentrations (high nM to ~ μM) of the enzyme. However, at 50% glycerol concentration the pseudo second-order rate constant for the hydrolysis of peptide II was determined to be 205 ± 20 M^{-1} s^{-1}, this value was comparable to the reported value of 50 to 104 M^{-1} s^{-1} (21,24,26) for peptides mimicking the same cleavage site. Shimizu *et al.* (15) have reported detection of no enzymatic activity with a full-length MBP-NS3 protease on peptide substrates representing the NS4A/4B cleavage site substrate and only a marginal second-order rate constant of 6 M^{-1} s^{-1} for a peptide (GDDIVPCSMSYTWI) representing the NS5A/5B cleavage site. Recently k_{cat}/K_M values ranging from 60 to 700 M^{-1} s^{-1} for NS3 protease domain and MBP-NS3 full-length protease were reported (28,32) for the NS5A/NS5B cleavage site.

Preformed complex of NS3 protease and NS4A peptide. It has been reported that the hydrophobic central core of the NS4A protein activates the NS3 protease activity in peptide cleavage assays (15,22,25,27,28). The effect of the 14-amino-acid peptide corresponding to residues 21 to 34 of the NS4A protein (4Apep21-34) on the activity of the NS3 protease has also been investigated with our detergent-free preparation of NS3. In this study, the 4Apep21-34 peptide was modified with the addition of two lysine residues to its C-terminus. The resultant peptide, 4Apep21-34KK, showed enhanced solubility. It was crucial for efficient activation by the 4Apep(21–34)KK to *preform* the complex at high concentration as described under Materials and Methods. For determining the kinetic parameters of the complex containing NS3 and a NS4A peptide, it was important that high concentrations of both the protease and activator were maintained until the complex was ready for assays so as not to allow dissociation of the components during the course of the reaction and that steady-state kinetic conditions were strictly observed.

For the determination of initial velocity, the enzymatic reaction was first followed for a period of 20 min at the lowest substrate concentration desired for substrate saturation. Formation of product as a function of time was found to be linear for a length of only 4 min with less than 10% substrate hydrolyzed (data not shown). Thus, all initial velocity data were obtained (see Materials and Methods) for reactions conducted for a period of less than 4 min. Figure 2 shows the initial velocity for peptide hydrolysis, catalyzed by the detergent-free preparation of NS3, as a function of substrate concentration for peptide II in the presence

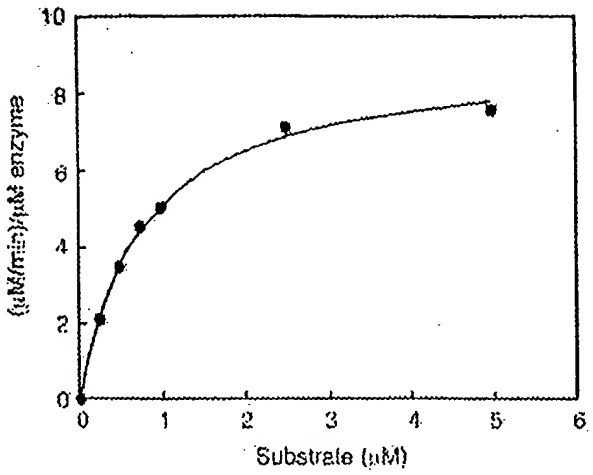


FIG. 2. Substrate saturation and initial velocity of peptide hydrolysis as catalyzed by the NS3 protease under steady-state kinetic conditions (see text). The enzyme reaction was carried out in 50 mM Hepes (pH 7.5), 10 mM DTT, and 50% glycerol in the presence of 0.25–5 μM peptide II (see Results), 2 nM NS3, and 1.25 μM 4Apep(21-34)KK. The reaction was allowed to continue for 2.5 min within which less than 10% of the substrate (at the lowest concentration assayed) was hydrolyzed. The data points shown were averages of 5 independent measurements at each substrate concentration. The SE (<5%) were smaller than the symbols used in the figure.

4Apep(21–34)KK. The observed k_{cat} and K_M values were 9 min^{-1} and 0.76 μM, respectively, resulting in a pseudo second-order rate constant (k_{cat}/K_M) of 196,000 M^{-1} s^{-1} an almost 1000-fold activation by the NS4A peptide, and at least 130-fold greater than those reported for a similar substrate (22,25,27).

To rule out the possibility that kinetic parameters obtained for the detergent-free NS3 for peptide hydrolysis were in any way affected by the coumarin molety at the N-terminus of the substrate, peptide I was used as a control substrate. In identical assays substituting peptide I for peptide II, the k_{cat} and K_M values observed were 8.6 min^{-1} and 2.8 μM, respectively, yielding a k_{cat}/K_M value of 51,000 M^{-1} s^{-1} which was significantly greater than the reported values of 90 to 1600 M^{-1} s^{-1} for the same cleavage site substrates (22,25,27,30,31). Thus, the high pseudo second-order rate constant observed for catalysis by the NS3 protease was not limited to peptides containing a coumarin molety. To further show that the detergent-free enzyme in general forms a very active complex with the NS4A cofactor and that its high catalytic efficiency was independent of substrate, a peptide mimicking the NS4B/5A cleavage site, 7-methoxycoumarin-4-acetyl-EDASTPCSGS-Nph-L, was used as a substrate. Values of 13 min^{-1} and 5180 M^{-1} s^{-1} were obtained for k_{cat} and k_{cat}/K_M. The latter may be compared to the reported values of 80 M^{-1} s^{-1} for a similar substrate (22). In presence of NS4A peptide (20-mer) k_{cat}/K_M values ranging from 2670 to 20,000 M^{-1} s^{-1} were recently reported for the NS5A/5B cleavage site (28,30,31,32). In our hands NS5A/5B substrates were not sufficiently stable to allow us to accurately determine the kinetic parameters. The steady-state kinetic parameters obtained in this study with the detergent-free NS3 protease are summarized in Table 2.

Stabilization of NS3 protease domain by 4A pep (21–34)KK cofactor. It is possible that the NS4A protein may stabilize the NS3 protease *in vivo* by protecting it from degradation by cellular enzymes (33). We examined the stability of the NS3 protease domain in the presence of NS4A peptide (21–34)KK peptide. The enzyme (100 nM) was incubated at 25°C in the presence and absence of 4A pep(21–34)KK (25 μM) and activity was followed for change in activity with increasing incubation time as shown in Fig. 3. While there was no

TABLE 2

Kinetic Parameters of HCV NS3 Protease Domain with Peptide Substrates[a]

Peptide substrate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})
(I) NS4A-NS4B site mimicking peptides			
Mca-DEMEEC*ASHLPYK-(ε-NHCOCH$_3$)[b]	0.36 ± 0.07	29.2 ± 3.4	205 ± 20
Mca-DEMEEC*ASHLPYK-(ε-NHCOCH$_3$)[c] + 4APep(21-34)KK	9.0 ± 0.02	0.76 ± 0.02	196,000 ± 9,000
Ac-DEMEEC*ASHLPYK-(ε-NHCOCH$_3$)[d] + 4APep(21-34)KK	8.6 ± 0.3	2.8 ± 0.3	51,000 ± 3,000
(II) NS4B–NS5A site mimicking peptide			
Mca-EDASTPC*SGSNphL[e] + 4APep(21-34)KK	13.6 ± 0.7	43.6 ± 7.9	5180 ± 670

[a] Kinetic parameters were determined as described under Materials and Methods in 50 mM Hepes (pH 7.5), 10 mM DTT and 50% glycerol. Data are mean values from at least three independent experiments.

Mca, 7-methoxycoumarin-4-acetyl.

[c] The assay was performed with 2 nM enzyme and 1.25 μM 4Apep(21-34)KK in 50 mM Hepes (pH 7.5), 10 mM DTT, and 50% glycerol buffer.

[d] The assay was carried out with 10 nM enzyme and 5 μM 4Apep(21-34) KK in 50 mM Hepes (pH 7.5), 10 mM DTT, and 50% glycerol buffer.

[e] The assay was conducted with 5 nM enzyme and 10 μM 4Apep(21-34)KK in 50 mM Hepes (pH 7.5), 10 mM DTT, and 50% glycerol buffer.

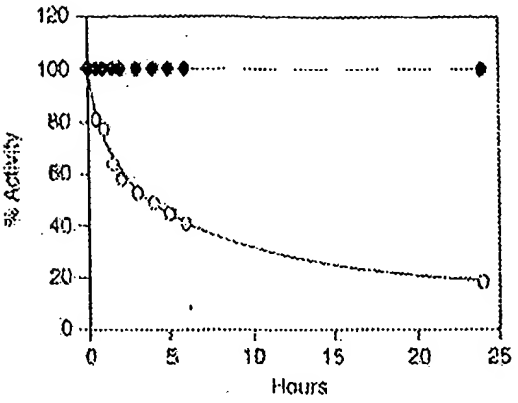


FIG. 3. Activity of NS3 protease (○) and NS3 protease/4Apep(21-34)KK complex (●) as a function of time. NS3 protease and NS3 protease/4Apep(21-34)KK complex was incubated in 50 mM Hepes (pH 7.5), 10 mM DTT, and 50% glycerol at 25°C and activity was determined by substrate cleavage assay at various time intervals. Activity at time zero was taken as 100% activity.

loss in activity observed in the presence of NS4A peptide, only 20% of the NS3 protease activity remained after 24 h at 25°C, in incubations without the NS4A peptide. Even lower concentrations of the enzyme (2 nM) showed no detectable loss of activity in presence of 4A pep(21–34)KK for at least 24 h at 25°C (data not shown).

Complex of the NS3 protease and 4A cofactor. Our kinetic data obtained for the NS3 protease in the presence of the NS4A cofactor suggest that the affinity of NS4A peptide to the NS3 protease domain may be very different under our assay conditions from those reported previously. It has been reported that maximal activity is achieved at 1:1 molar ratio of NS3 protease domain and the NS4A peptide (22). In the absence of detergents, a 10- to 14-fold excess of NS4A peptide has been used to attain a maximal activation of the MBP–NS3 fusion protein ($k_{cat}/K_M \approx 250\ M^{-1}\ s^{-1}$) and the protease domain ($k_{cat}/K_M \approx 2670\ M^{-1}\ s^{-1}$) in catalysis of hydrolysis of NS5A/5B cleavage-site-mimicking substrates (15,28). In the present study, we have demonstrated that under detergent-free conditions the NS3 protease domain interacts with the NS4A peptide to yield a significantly more active catalytic complex, although at a significantly higher NS4Apep(21–34)KK to NS3 ratio of 625, due to a moderate K_d of ≈20 μM of the cofactor to achieve complete saturation of the protease. We note that the values of the rate constants (k_{cat}, k_{cat}/K_M) shown in Table 2 are likely underestimates because they are not obtained in the presence of theoretical saturating amounts of 4Apep(21–34)KK. In 20% glycerol-containing buffers, the affinity of NS4A peptide to NS3 protease domain is weaker than that in solution containing 50% glycerol (37), and complete saturation is not attainable due to the limited solubility of the NS4A peptide (150 μM in assay buffer). The high activity of the complex correlates well the recently published three-dimensional structures of the complexed and uncomplexed form of the NS3 protease domain (34–36). In the presence of the 4A peptide the NS3 protease catalytic triad forms a chymotrypsin-like fold, similar to other serine proteases, with the optimal orientation of the active site Asp, His, and Ser residues, which is not properly formed in the uncomplexed form (35). Also the stabilization effect of 4A peptide is evident by the interactions observed with the N-terminus of the protease. The 4A peptide beta sheet becomes part of the N-terminal domain beta sheet of the protease and stabilizes the structure. We and others (22,37) have found that 4A binding as well as the proteolytic activity of the complex increases with increasing concentrations of glycerol, with maximal activity at ≈50% glycerol. *In vivo*, the interaction between NS4A and NS3 is likely to be substantially more complicated since the helicase domain (29,38–42), encoded downstream and in frame with the protease as a single NS3 polypeptide, is likely to be part of an integral protease–helicase–activator multienzyme complex and plays a critical role in regulating virus replication.

CONCLUSION

A rapid purification procedure in the absence of detergents has been devised for the NS3 protease domain (residues 1027 to 1206) of the human hepatitis C virus, and an efficient *in vitro* assay is described to demonstrate that detergents are not required for the purification, solubility, stability, and for the optimal activity of the recombinant NS3 protease. The activity of the HCV NS3 protease is comparable to those reported for other viral proteases such as the HIV-1 protease (43), adenovirus protease (44), cytomegalovirus protease (45,46), and herpes simplex type-1 protease (47). The highly active noncovalent complex of the N-terminus domain of NS3 protease and NS4A peptide along with the more stable NS4A/4B cleavage site-mimicking peptide substrate will be useful tools for screening inhibitors as therapeutic leads for HCV therapy.

ACKNOWLEDGMENTS

We are grateful to Drs. R. Cortese and R. De Francesco for providing us with the plasmid of the NS3 protease domain. We thank Mr. Brett Johns for technical assistance, Mr. Jimmy Ramos Calaycay for mass spectrometry, and Ms. Theresa Wood for amino acid sequence analysis.

REFERENCES

1. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., and Houghton, M. (1989) Isolation of a cDNA clone derived from

a blood-borne non-A non-B viral hepatitis genome. *Science* **244**, 359–362.

2. Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tagtmeier, G. E., Bonino, F., Colombo, M., Lee, W. S., Kuo, C., Berger, K., Shister, J. R., Overby, L. R., Bradley, D. W., and Houghton, M. (1989) An assay for circulating antibodies to a major etiologic virus of human non-A non-B hepatitis. *Science* **244**, 362–364.
3. Chien, D., Choo, Q. L., Tabrizi, A., Kuo, C., McFarland, J., Berger, K., Lee, C., Shuster, J., Nguyen, T., Moyer, D., Tong, M., M., Furuta, S., Omata, M., Tegtmeyer, G., Alter, H., Schiff, E., Jeffers, L., Houghton, M., and Kuo, G. (1992). Diagnosis of hepatitis C virus (HCV) infection using an immunodominant chimeric polyprotein to capture circulating antibodies: Reevaluation of the role of HCV in liver disease. *Proc. Natl. Acad. Sci. USA* **89**, 10011–10015.
4. Choo, Q. L., Richman, K. H., Ham, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Colt, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G., and Houghton, M. (1991) Genetic organization and diversity of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
5. Kato, M., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K. (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc. Natl. Acad. Sci.* USA **87**, 9524–9528.
6. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onoshi, E., Andoh, T., Yoshida, I., and Okayama, H. (1991) Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol.* **65**, 1105–1113.
7. Grakoui, A., Wychowski, C., Lin, C., Feinstone, S. M., and Rice, C. M. (1993) Expression and identification of hepatitis C virus polyprotein cleavage products. *J. Virol.* **67**, 1385–1395.
8. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohno, K. (1991) Gene mapping of the putative structural region of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**, 5547–5551.
9. Tomei, L., Failla, C., Santolini, E., De Francesco, R., and La Monica, N. (1993) NS3 is a serine protease required for processing hepatitis C virus polyprotein. *J. Virol.* **67**, 4017–4026.
10. Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1993) Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J. Virol.* **67**, 3835–3844.
11. Failla, C., Tomei, L., and De Francesco, R., (1995) An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. *J. Virol.* **69**, 1769–1777.
12. Lin, C., Pragai, B. M., Grakoui, A., Xu, J., and Rice, C. M. (1994) Hepatitis C virus NS3 serine proteinase: *trans*-cleavage requirements and processing kinetics. *J. Virol.* **68**, 8147–8157.
13. Failla, C., Tomei, L., and De Francesco, R. (1994) Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J. Virol.* **68**, 3753–3760.
14. Lin, C., Thomson, J. A., and Rice, C. M. (1995) A central region in the hepatitis C virus NS4A protein allows formation of an active NS3–NS4A serine proteinase complex *in vivo* and *in vitro*. *J. Virol.* **69**, 4373–4380.

    Shimizu, Y., Kamaji, K., Masuho, Y., Yokota, T., Inoue, H., Sudo, K., Satoh, S., and Shimotohno, K. (1996) Identification of the sequence on NS4A required for enhanced cleavage of the NS5A/5B site by hepatitis C virus NS3 protease. *J. Virol.* **70**, 127–132.
16. Lin, C., and Rice, C. M. (1995) The hepatitis C virus NS3 serine proteinase and NS4A cofactor: Establishment of a cell-free *trans*-processing assay. *Proc. Natl. Acad. Sci. USA* **92**, 7622–7626.
17. Tomei, L., Failla, C., Vitale, R. L., Bianchi, E., and De Francesco, R. (1996) A central hydrophobic domain of the hepatitis C virus NS4A protein is necessary and sufficient for the activation of the NS3 protease. *J. Gen. Virol.* **77**, 1065–1070.
18. Bartenschlager, R., Lohmann, V., Wilkinson, T., and Koch, J. O., (1995) Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for the polyprotein maturation. *J. Virol.* **69**, 7519–7528.
19. Butkiewicz, N. J., Wendel, M., Zhang, R., Jubin, R., Pichardo, J., Smith, E. B., Hart, A. M., Ingram, R., Durkin, J., Mui, P. W., Murray, M. G., Ramanathan, L., and Dasmahapatra, B. (1996) Enhancement of hepatitis C virus NS3 proteinase activity by association with NS4A-specific synthetic peptides—Identification of sequence and critical residues of NS4A for the cofactor activity. *Virology* **225**, 328–338.
20. D'Souza, E., D., A., Grace, K., Sangar, D. V., Rowlands, D. J., and Clarke, B. E. (1995) *In-vitro* cleavage of hepatitis C virus polyprotein substrates by purified recombinant NS3 protease. *J. Gen. Virol.* **76**, 1729–1736.
21. Kakiuchi, N., Hijikata, M., Komoda, Y., Tanji, Y., Hirowatari, Y., and Shimotohno, K. (1995) Bacterial expression and analysis of cleavage activity of HCV serine proteinase using recombinant and synthetic substrates. *Biochem. Biophys. Res. Commun.* **210**, 1059–1065.
22. Steinkuhler, C., Urbani, A., Tomei, L., Sardana, M., Bianchi, E., Pessi, A., and De Francesco, R. (1996) Activity of purified hepatitis C virus protease NS3 on peptide substrates. *J. Virol.* **70**, 6694–6700.
23. Shoji, I., Suzuki, T., Chieda, S., Sato, M., Harada, T., Chiba, T., Matsuura, Y., and Miyamura, T. (1995) Proteolytic activity of NS3 serine protease of hepatitis C virus effeciently expressed in *Escherichia coli*. *Hepatology* **22**, 1648–1655.
24. Mori, A., Yamada, K., Kimura, J. Kolde, T., Yuasa, S., Yamada, E., and Miyamura, T. (1996) The N-terminal region of NS3 serine proteinase of hepatitis C virus is important to maintain its enzymatic integrity. *FEBS Lett.* **378**, 37–42.
25. Steinkuhler, C., Tomei, L., and De Francesco, R., (1996) *In-vitro* activity of hepatitis C virus protease NS3 purified from recombinant baculovirus-infected sf9 cells. *J. Biol. Chem.* **271**, 6367–6373
26. Suzuki, T., Sato, M., Chieda, S., Shoji, I., Harada, T., Yamadawa, Y., Watabe, S., Matsuura, Y., and Miyamura, T. (1995) In-vivo and in-vitro transcleavage activity of hepatitis C virus proteinase expressed by recombinant baculoviruses. *J. Gen. Virol.* **76**, 3021–3029.
27. Urbani, A., Bianchi, E., Narjes, F., Tramontano, A., De Francesco, D., Steinkuhler, C., and Pessi, A. (1997) Substrate specificity of the hepatitis C virus serine protease NS3. *J. Biol. Chem.* **272**, 9204–9209.
28. Vishnuvardhan, D., Kakiuchi, N., Urvil, P. T., Shimotohno, K., Kumar, P. K. R., and Nishikawa, S. (1997) Expression of highly active recombinant NS3 protease domain of hepatitis C virus in *E. coli*. *FEBS Lett.* **402**, 209–212.
29. Hong, Z., Ferrari, E., Wright-Minogue, J., Chase, R., Risano, C., Seelig, G., Lee, C. G., and Kwong, A. D. (1996) Enzymatic characterization of hepatitis C virus NS3/4A complexes expressed in mammalian cells by using the herpes simplex virus amplicon system. *J. Virol.* **70**, 4261–4268.
30. Landro, J. A., Raybuck, S. A., Luong, Y. P. C., O'Malley, E. T., Harbeson, S. L., Morgenstern, K. A., Rao, G., and Livingston,

D. J. (1997) Mechanistic role of an NS4A peptide cofactor with the truncated NS3 protease of hepatitis C virus: Elucidation of the NS4A stimulatory effect via kinetic analysis and inhibitor mapping. *Biochemistry* **36**, 9340–9348.

31. Zhang, R., Durkin, J., Windsor, W. T., McNemar, C., Ramanathan, L., and Le, H. U. (1997) Probing the substrate specificity of hepatitis C virus NS3 serine protease by using synthetic peptides. *J. Virol.* **71**, 6208–6213.
32. Kwong, A. D., Kim, J. L., Rao, G., Lipovsek, D., and Raybuck, S. A. (1998) Hepatitis C virus NS3/NS4A protease. *Antiviral Res.* **40**, 1–18.
33. Tanji, Y., Hijikata, M., Satoh, S., Kaneko, T., and Shimotohno, K. (1995) Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *J. Virol.* **69**, 1575–1581.
34. Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chamber, S. P., Markland, W., Lapre, C. A., O'Melley, E. T., Harbeson, S. L., Rice, C. M., Murco, M. A., Caron, P. R., and Thomson, J. A. (1996) Crystal structure of the hepatitis C Virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **87**, 343–355.
35. Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., and Hostomska, Z. (1996) The crystal structure of the hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* **87**, 331–342.
36. Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J., Sardana, M. K., Steinkuhler, C., Tomei, L., DeFrancesco, R., Kuo, L. C., and Chen, Z. (1998) Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: A 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci.* **7**, 837–847.
37. Bianchi, E., Urbani, A., Biasiol, G., Brunetti, M., Pessi, A., Defrancesco, R., and Steinkuhler, C. (1997) Complex formation between the hepatitis C virus serine protease and a synthetic NS4A cofactor peptide. *Biochemistry* **36**, 7890–7897.
38. Kim, D. W., Gwack, Y., Han, J. H., and Choe, J. (1995) C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biohem. Biophys. Res. Commun.* **215**, 160–166
39. Jin, L., and Peterson, D. L. (1995) Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA helicase. *Arch. Biochem. Biophys.* **323**, 47–53
40. Kanai, A., Tanabe, K., and Kohara, M., (1995) Poly(U) binding activity of hepatitis C virus NS3 protein, a putative RNA helicase. *FEBS Lett.* **376**, 221–224
41. Suzich, J. A., Tamura, J. K., Palmer-Hill, F., Warrener, P., Grakoui, A., Rice, C. M., Feinstine, S. M., and Collett, M. S., (1993) Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparision with the related pestivirus and flavivirus enzymes. *J. Virol.* **67**, 6152–6158.
42. Morgenstern, K. A., Landro, J. A., Hsiao, K., Lin, C., Gu, Y., Su, M. S.-S., and Thomson, J. A. (1997) Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3–NS4A complex isolated from transfected cos cells. *J. Virol.* **71**, 3767–3775.
43. Sardana, V. V., Schlabach, A. J., Graham, P., Bush, B. L., Condra, J. H., Culberson, J. C., Gotlib, L., Graham, D. J., Kohl, N. E., LaFemina, R. L., Schneider, C. L., Wolanski, B. S., Wolfgang, J. A., and Emini, E. A. (1994) *Biochemistry* **33**, 2004–2010.
44. Mangel, W. F., Toledo, D. L., Brown, M. T., Martin, J. H., and McGrath, W. J. (1996) Characterization of three components of human adenovirus proteinase activity in-vitro. *J. Biol. Chem.* **271**, 536–543.
45. Sardana, V. V., Wolfgang, J. A., Veloski, C. A., Long, W. J., LeGrow, K., Wolanski, B., Emini, E. A., and LaFemina, R. L. (1994) Peptide substrate cleavage specificity of the human cytomegalovirus protease. *J. Biol. Chem.* **259**, 14337–14340.
46. Darke, P. L., Chen, E., Hall, D. L., Sardana, M. K., Veloski, C. A., LaFemina, R. L., Shafer, J. A., and Kuo, L. C. (1994) Purification of active herpes simplex virus-1 protease expressed in *Escherichia coli*. *J. Biol. Chem.* **269**, 18708–18711.
47. Hall, D. L., and Darke, P. L. (1995) Activation of the herpes simplex virus type-1 protease. *J. Biol. Chem.* **270**, 22697–22700.

FEBS 18138

FEBS Letters 402 (1997) 209–212

Expression of highly active recombinant NS3 protease domain of hepatitis C virus in *E. coli*

Daesety Vishnuvardhan[a], Nobuko Kakiuchi[b], Petri T. Urvil[a], Kunitada Shimotohno[b], P.K.R. Kumar[a,*], Satoshi Nishikawa[a]

[a] *National Institute of Bioscience and Human Technology, AIST, 1-1 Higashi, Tsukuba, Ibaraki 305, Japan*
[b] *Virology Division, National Cancer Center Research Institute, Tsukiji, Tokyo 104, Japan*

Received 22 November 1996

Abstract The serine protease domain of HCV comprising amino acids 1027–1218 (ΔNS3) was expressed in *E. coli* with a His tag at its N-terminal end. The protease was purified to apparent homogeneity by a single step affinity chromatography resulting in high yields (~3 mg/l of cultured cells). The ΔNS3 efficiently cleaves a 17-mer peptide corresponding to the NS5A-NS5B junction with $k_{cat}/K_m = 160 \times 10^{-3}$ min^{-1} μM^{-1} in the presence of NS4A peptide. Our ΔNS3 represents the minimal domain possessing highly active protease of NS3 constructed so far. The ΔNS3 protein also efficiently processed a longer substrate corresponding to NS5A/5B junction (2203–2506 amino acids) that was synthesized by in vitro transcription and translation system.

Key words: Hepatitis C virus; Serine protease; Non-structural protein 3

1. Introduction

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion non-A, non-B hepatitis, is an enveloped virus containing a single-stranded RNA genome of approximately 9.5 kb nucleotides [1,2]. A single polyprotein of 3010–3030 amino acids is translated from this genome [3] in the order of NH2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH (Fig. 1A) [4,5]. This polyprotein is subsequently processed by a combination of host and viral proteases to produce at least 10 viral proteins. The core protein (C) and envelope proteins (E1 and E2) are structural proteins, and NSs are non-structural proteins [6,7].

Previous studies indicate that a host signal peptidase localized in endoplasmic reticulum (ER) catalyzes polyprotein cleavages in the structural region (C/E1, E1/E2, E2/p7 and p7/NS2) [8,5], whereas a HCV encoded serine protease located in the N-terminal one-third of the NS3 protein is responsible for cleavages at four sites (3/4A, 4A/4B, 4B/5A and 5A/5B) (Fig. 1A) [9–14] in the NS region. Using a transient coexpression system it has been shown that proteolytic cleavages in the non-structural protein region (NS3 to NS5B) of HCV polyprotein are effected by two viral proteins, NS3 and NS4A [15–19]. The N-terminal 180 amino acid region of NS3 includes sequences showing homology with the active sites of serine proteases [20–22]. Histidine 1083, aspartate 1107 and serine 1165 (numbers are according to their locations in the polyprotein of HCV subtype J (HCV-1b) [23,24]) found in this domain have been proposed to constitute the catalytic triad of the NS3 protease similar to other serine proteases belonging to the chymotrypsin family.

NS4A is shown to be the NS3 protease cofactor or effector enhancing cleavage efficiency at various sites NS3/4A, NS4A/4B, NS4B/5A and NS5A/5B [16,19,25–27]. NS4A is an amphipathic protein of 54 amino acids and it has a very hydrophobic N-terminal domain followed by a hydrophilic C-terminal domain [16]. Using an in vitro reconstituted assay system it has been shown that the residues 22–31 of NS4A constitute the putative core sequence for NS4A's effector activity [27]. The mechanism by which NS4A facilitates cleavage remains obscure. Truncation experiments have mapped the N-terminus of NS3 as the domain responsible for interaction with NS4A [25,27].

Since the NS3 protein is very important for releasing functional proteins from the polyprotein, it is currently being targeted in the development of drugs and diagnostics. As a matter of fact, several known serine protease inhibitors have been tested for their action on NS3 activity in vitro and it was found that millimolar concentrations of these compounds were required to show moderate inhibitory effect on NS3 protease activity [3,28]. One major obstacle for these well known inhibitors may be the presence of helicase domain also or alternatively NS3 protease may be structurally different in comparison with other serine protease domains.

In order to carry out detailed characterization of this enzyme in terms of its substrate specificity, kinetics, sensitivity to inhibitors and for structural studies, a reproducible and convenient large-scale purification of the enzymatically active protease domain of NS3 is essential. In this report we show that the region encompassing amino acids 1027–1218 tagged with the His tail (6 His) has greater activity than other expression constructs made so far and the purification procedure is much simpler. We also show that the purified protein possesses the ability to interact with NS4A resulting in increased proteolytic activity.

2. Materials and methods

2.1. Construction of expression plasmid containing the HCV serine protease domain

To construct the expression plasmid pHisBΔNS3, a cDNA fragment encoding amino acid residues 1027–1218 in the HCV polyprotein was obtained by PCR using appropriate oligonucleotides (1: d(CCGCTGCAGCCATGGCGCCTATCACGGCCTAT), 2: d(CCGAAGCTTTCAGGCCGGAGGGGATGAGTT)), which insert a *Pst*I site at the 5'-end and a TAG (stop) codon and *Hin*dIII site at the 3'-end of the sequence. We used plasmid pMANS34NSH [29] as a template to amplify amino acid residues 1027–1218 in the HCV polyprotein. The PCR product amplified using Ex. Taq (Takara) was first cloned into pCR II vector (Invitrogen) according to the manufactur-

*Corresponding author. Fax: (81) (298) 54-6095.

PII S0014-5793(96)01532-3

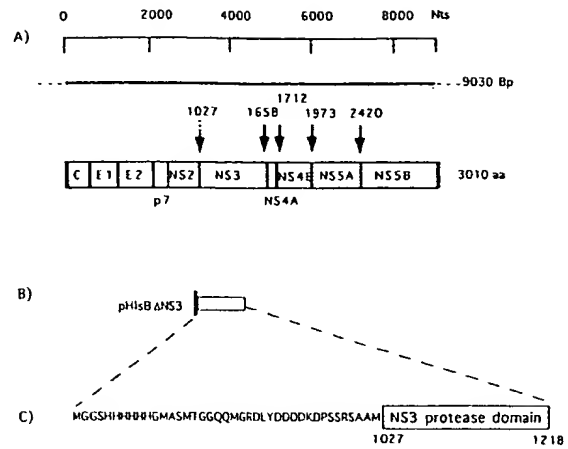


Fig. 1. Schematic representation of HCV serine proteinase domain expressed in *E. coli*. A: The HCV genome and a translated product with names of the processed proteins. The filled arrow shows the cleavage site of HCV serine protease, whereas the dotted arrow indicates the cleavage site of HCV metaloprotease. B: ΔNS3. Closed and open boxes indicate the 6 histidine tag and the protease domain, respectively. C: N-terminal extra amino acid sequence and NS3 protease domain.

er's instructions and digested with *Pst*I and *Hin*dIII to release the fragment encompassing amino acids 1027–1218 of HCV polyprotein. It was subsequently cloned into expression plasmid pTrcHisB (Invitrogen) which was also digested with *Pst*I and *Hin*dIII. The resulting plasmid pHisBΔNS3 (Fig. 1B,C) encodes the protease domain of NS3 (192 amino acids) with a N-terminal 40 non-virus encoded amino acids possessing a consecutive stretch of 6 His residues that allows fusion protein to be purified in a single step by metal chelating affinity chromatography. The cloned DNA fragment was sequenced in order to exclude the introduction of mutations by PCR and also to confirm the in-frameness of the insert.

2.2. Expression and purification of HCV NS3 protease domain

Following transformation of *E. coli* HB101, cells harboring the vector pHisBΔNS3 were grown in LB medium containing 100 μg/ml ampicillin. When the absorbance reached a value of 0.5–0.6 OD_{600} isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to give a final concentration of 1 mM and the incubation was continued for an additional 2 h at 37°C. Under this induction condition a high level expression of ΔNS3 was observed and there was only a small amount present in the insoluble fraction. The cells were harvested by centrifugation and washed extensively with PBS (20 mM sodium phosphate; pH 7.4, 140 mM NaCl). The cell pellet was resuspended in lysis buffer (20 mM sodium phosphate; pH 6.3, 500 mM NaCl) and disrupted by sonication on ice using a Branson 200 sonifier (30 s × 6 strokes at 18 W output with 30 s intervals). The homogenate was centrifuged at 30 000 × *g* for 30 min to remove cell debris and was chromatographed on a nickel-agarose column (Qiagen). We applied a stepwise gradient of pH to elute the protease in order to overcome the precipitation of the enzyme. The column was washed extensively with several column volumes of lysis buffer and subsequently washed using the same buffer of pH 6.0 and finally with the same buffer of pH 5.0. Resin bound protein was eluted with sodium phosphate buffer (pH 4.0) containing 500 mM NaCl. Eluted fractions were subjected to SDS-PAGE [30], protein containing fractions were pooled and concentrated by using Millipore Ultrafree Biomax 10K concentrator (Millipore). The enzyme was stored in aliquots at −20°C in the same buffer containing 40% glycerol.

Protein concentrations were estimated from UV absorbance at 280 nm: an extinction coefficient of ε = 20 800 M^{-1} cm^{-1} was calculated on the basis of primary sequence data according to published procedures [31] and concentration of ΔNS3 was determined according to the Lambert Beer law. Alternatively, ΔNS3 concentration was determined by Bradford (BioRad) assay using lysozyme as standard.

2.3. HCV NS3 protease assays

To characterize the enzymatic activity of the purified protease (ΔNS3) we investigated its ability to cleave a synthetic peptide, S-1 (Dns-Gly-Glu-Ala-Gly-Asp-Asp-Ile-Val-Pro-Cys-Δ-Ser-Met-Ser-Tyr-Thr-Trp-Thr-COOH, Δ; cleavage site) corresponding to the cleavage site of the NS5A/5B. Protease activity of ΔNS3 and maltose binding NS3 fusion protein, MBP-NS3 (amino acids 985–1647) [29] was analyzed either in the presence or absence of NS4A peptide, P41 (amino acids 1673–1692 of HCV polyprotein).

Kinetic constants were determined from enzyme assays with the synthetic substrate concentration ranging from 25 to 960 μM. The K_m and k_{cat} values were determined by Lineweaver-Burk plots. Kinetic reactions were analyzed either in the presence or absence of P41. For evaluating the rate of reaction in the absence of P41, the ΔNS3 or MBP-NS3 (0.72 mM) was incubated with various concentrations of substrate in a buffer (Tris-HCl (pH 7.8), 30 mM $NaCl_2$, 5 mM $CaCl_2$, 10 mM DTT) at 25°C for 60 min, whereas reactions carried out in the presence of P41 (10 mM) were initially equilibrated with ΔNS3 or MBP-NS3 for 15 min at 37°C in the above buffer. The reaction was initiated by addition of substrate at 37°C and data points were collected for 10 min.

In order to test the ability of the ΔNS3 to cleave longer substrate representing NS5A/B junction protein in the amino acid sequence 2203–2506 of HCV polyprotein which closely resembles the situation in vivo, radiolabelled NS5A/5B substrate was produced using a coupled transcription-translation system (TNT Promega) according to the manufacturer's protocol. The DNA encoding amino acids 2203–2506 of polyprotein representing the NS5A/B site was amplified by PCR and was added to a 25 μl TNT reaction in the presence of [^{35}S]methionine (Amersham) at 300 μCi/ml and incubated at 30°C for 1–2 h. For experiments to assess proteolytic cleavage of pre-formed substrate the TNT reaction mixture was diluted by adding an equal volume of buffer (100 mM HEPES (pH 7.6), 300 mM NaCl, 6 mM $MgCl_2$, 20 mM DTT). To 6 μl (≈ 300 cpm of translated product) of diluted TNT reaction mixture, ΔNS3 (750 nM) was added and incubated for 30 min at 30°C. In experiments to study the effect of NS4A peptide, the protease was pre-incubated with P41 for 5 min at 30°C before the addition of substrate. Samples were withdrawn at different time intervals and the reaction was stopped by adding SDS-PAGE sample buffer followed by denaturing at 95°C for 3 min. All samples were loaded on to SDS-PAGE and the radioactivity was quantitated by image analyzer (BAS 2000, Fuji Film).

3. Results and discussion

The purified ΔNS3 protease migrated as a single band with a molecular mass consistent with that calculated from the primary sequence (molecular mass of 25 kDa as judged from SDS-PAGE) (Fig. 2). The final yield was ~3 mg of purified protein per liter of cultured cells and the purity of the preparate was estimated to be over 95% as judged by SDS gel. Though several constructs have been reported for the expression of the NS3, considering the yield and the simplicity of the purification procedure which we described here offers clear advantage for structural studies, wherein large amounts of enzyme are required. Table 1 shows that in the absence of P41 the K_m of ΔNS3 was 250 μM and k_{cat} was 2.0 min^{-1}. In

Table 1
Effect of P41 on the cleavage kinetics of peptide NS5A/5B by ΔNS3 and MBP-NS3[a]

Enzyme	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1} μM^{-1})
ΔNS3	250	2.0	8×10^{-3}
ΔNS3+P41	99	15.8	160×10^{-3}
MBP-NS3	360	1.3	3.6×10^{-3}
MBP-NS3+P41	196	5.1	26×10^{-3}

[a]Data are the mean values from three independent experiments.

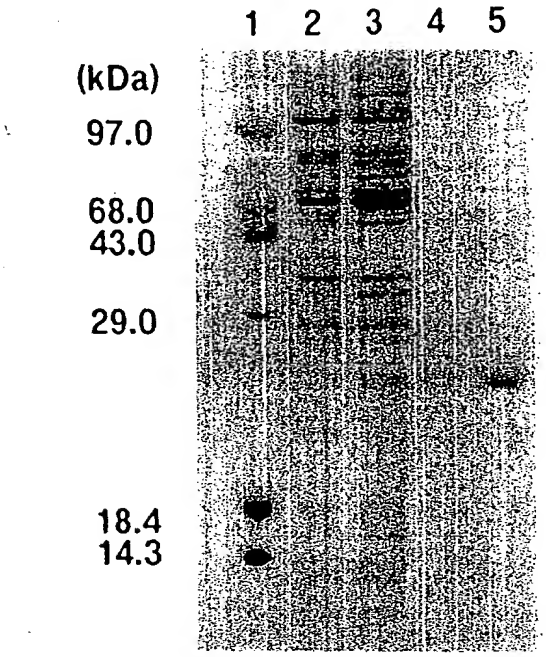


Fig. 2. Purification of the NS3 protease domain. Samples deriving from single steps of the purification were loaded on an SDS-15% polyacrylamide gel, and bands were visualized by Coomassie staining. Lane 1: molecular mass markers; lane 2: homogenate from *E. coli* without construct; lane 3: homogenate from pTrcHisBΔNS3; lane 4: 750 ng of purified ΔNS3 after affinity chromatography; lane 5: 1.5 μg of purified ΔNS3 after affinity chromatography.

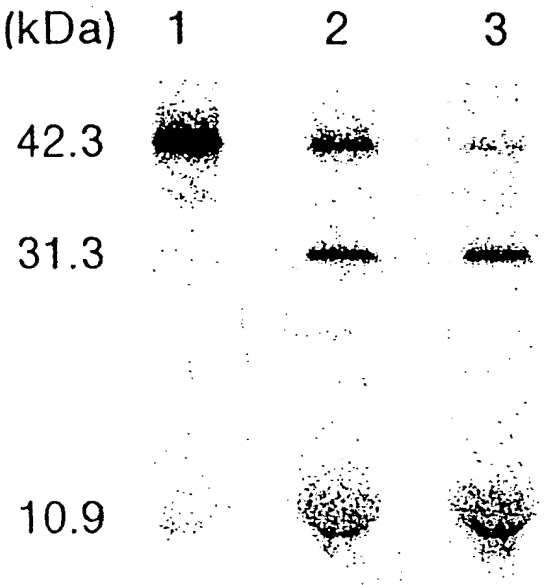


Fig. 3. TNT analysis of cleavage of NS5A/5B substrate by ΔNS3. Substrate representing NS5A/5B incubated, lane 1: without ΔNS3; lane 2: with ΔNS3; lane 3: with both ΔNS3 and P41. All samples were incubated at 30°C for 30 min; proteins were labelled with [^{35}S]methionine and analyzed by SDS-PAGE followed by fluorography.

contrast, in the presence of a 10-fold molar excess of P41 the K_m (99 μM) was about 2.5 times lower than that in the case of ΔNS3 alone, and the k_{cat} (15.8 min^{-1}) was about 8 times higher than that for ΔNS3 alone. As a result the k_{cat}/K_m value increased 20 times in the presence of P41. When we compared the k_{cat}/K_m value of MBP-NS3, ΔNS3 protease activity was increased to 2-fold in the absence and 6-fold in the presence of P41. Our ΔNS3 protease activity was found to be the highest reported so far. The data presented above suggest that region 1027–1218 represents the minimal domain required for the protease activity, since the region 1027–1218 exemplifies all the cleavage kinetics of the MBP-NS3 (985–1647) reported earlier [27].

The coefficient for proteolytic efficiency was comparable to that of two reports published recently [27,32], whereas the K_m and k_{cat} values are comparable with only that of NS3 (encompassing both the protease and helicase domains fused to MBP) [27]. In the case of ΔNS3, the presence of P41 clearly increases the affinity for the substrate (as evident from K_m values) and thereby increases the catalytic rate (Table 1). Shimizu et al. [27] have made a similar observation with NS3 which has both protease and helicase functional domains, whereas Steinkuhler et al. [32] have expressed NS3 protease domain encompassing amino acids 1038–1226 of HCV polyprotein and analyzed its activity on NS4A/NS4B cleavage site both in the presence and in the absence of NS4A protein. Though the kinetic differences were observed using a peptide derived from a NS4A-independent cleavage site (NS4A/4B), the data clearly show the activation role of NS4A in altering the NS3 activity in general. However, the affinity (K_m) was not altered in the presence of NS4A as evident from their published data. This may possibly be due to the absence of essential N-terminal amino acids for interaction with NS4A in their expressed protease domain. The fact that NS4A can interact with ΔNS3 in a similar way as with full NS3 and brings about similar changes confirms that ΔNS3 despite its truncated version and fusion retains the favorable conformation for interaction. Based on the above observations we can conclude that amino acid residues 1027–1218 are sufficient to represent the complete protease function of NS3 even considering its ability to interact with NS4A. Detailed analyses of

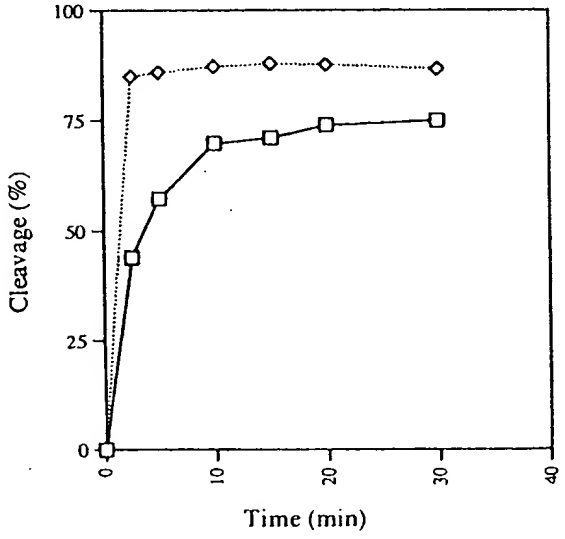


Fig. 4. TNT analysis on effect of P41 on ΔNS3 rate of cleavage. Normalized cleavage % is shown against different time intervals. Without P41 (□) and with P41 (◇).

the regulation and the substrate requirements of the protease are expected to offer further insights into the mechanism of activation by P41.

In addition, when we tested the ability of ΔNS3 to cleave a longer substrate representing the NS5A/B junction, as shown in Fig. 3, ΔNS3 could cleave the NS5A/B junction protein (amino acids 2203–2506) in close agreement with the observation when the small synthetic peptide substrate NS5A/B was used. The addition of P41 again clearly enhanced the cleavage reaction. From the time course of these reactions (Fig. 4) the rate of cleavage was stimulated more than 2-fold by the addition of P41. Since the purified recombinant ΔNS3 protease has a relatively smaller molecular size, representing the complete catalytic function of NS3 protease, and considering the ease and yield of purification, we believe that it is an ideal candidate to generate further studies elucidating the structure-function relationships, mechanism of interaction with NS4A peptide as well as developing protease inhibitors.

Acknowledgements: We would like to thank Dr. Sang Wang Gal and Mr. Keigo Machida for helpful discussions. This investigation was supported by grants from the New Energy and Industrial Technology Development Organization (NEDO) of Japan to S.N., P.K.R.K. and D.V.

References

[1] Choo, Q.-L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. and Houghton, M. (1989) Science 244, 359–362.
[2] Choo, Q.-L., Richman, K., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P., Weiner, A.J., Bradley, D.W., Kuo, G. and Houghton, M. (1991) Proc. Natl. Acad. Sci. USA 88, 2451–2455.
[3] Matsuura, Y. and Myamura, T. (1993) Semin. Virol. 4, 297–304.
[4] Grakoui, A., Wyochowski, C., Lin, C., Feinstone, S.M. and Rice, C.M. (1993) J. Virol. 67, 1385–1395.
[5] Lin, C., Lindenbach, B.D., Pragai, B., McCourt, D.W. and Rice, C.M. (1994) J. Virol. 68, 5063–5073.
[6] Houghton, M. (1996) in: Virology (Fields, B.N., Knipe, D.M. and Howley, P.M., Eds.) pp. 1035–1058, Raven Press, New York.
[7] Rice, C.M. (1996) in: Virology (Fields, B.N., Knipe, D.M. and Howley, P.M., Eds.) pp. 931–960, Raven Press, New York.
[8] Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M. and Shimotohno, K.(1991) Proc. Natl. Acad. Sci. USA 88, 5547–5551.
[9] Bartenschlager, R., Ahlborn-Laake, L., Mous, J. and Jacobsen, H. (1993) J. Virol. 67, 3835–3844.
[10] Eckart, M.R., Selby, M., Masiarz, F., Lee, C., Berger, K., Crawford, K., Kuo, C., Kuo, G., Houghton, M. and Choo, Q.-L. (1993) Biochem. Biophys. Res. Commun. 192, 399–406.
[11] Grakoui, A., McCourt, D.W., Wychowski, C., Feinstgone, S.M. and Rice, C.M. (1993) J. Virol. 67, 2832–2843.
[12] Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K. and Shimotohno, K. (1993) J. Virol. 67, 4665–4675.
[13] Manabe, S., Fuke, I., Tanishita, O., Kaji, C., Gomi, Y., Yoshida, S., Mori, C., Takamizawa, A., Yoshida, I. and Okayama, H. (1994) Virology 198, 636–644.
[14] Tomei, L., Failla, C., Santolini, E., DeFrancesco, R. and La-Monica, N. (1993) J. Virol. 67, 4017–4026.
[15] Bartenschlager, R.L., Ahlborn-Laake, L., Mous, J. and Jacobsen, H. (1994) J. Virol. 68, 5045–5055.
[16] Failla, C., Tomei, L. and De Francesco, R. (1994) J. Virol. 68, 3753–3760.
[17] Lin, C., Pragai, B.M., Grakoui, A., Xu, J. and Rice, C.M. (1994) J. Virol. 68, 8147–8157.
[18] Lin, C., Thomson, J.A. and Rice, C.M. (1995) J. Virol. 69, 4373–4380.
[19] Tanji, Y., Hijikata, M., Satoh, S., Kaneko, T. and Shimotohno, K. (1995) J. Virol. 69, 1575–1581.
[20] Kato, N., Hijikata, M., Nakagawa, M., Ootsuyama, Y., Muraiso, K., Ohkoshi, S. and Shimotohno, K. (1991) FEBS Lett. 280, 325–328.
[21] Pizzi, E., Tramontano, A., Tomei, L., La Monica, N., Failla, C., Sardana, M., Wood, T. and De Francesco, R. (1994) Proc. Natl. Acad. Sci. USA 91, 888–892.
[22] Rawlings, N.D. and Barrett, A.J. (1994) Methods Enzymol. 244, 19–61.
[23] Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. and Shimotohno, K. (1990) Proc. Natl. Acad. Sci. USA 87, 9524–9528.
[24] Tanaka, T., Kato, N., Nakagawa, M., Ootsuyama, Y., Cho, M.-J., Nakazawa, T., Hijikata, M., Ishimura, Y. and Shimotohno, K. (1992) Virus Res. 23, 39–53.
[25] Failla, C., Tomei, L. and DeFrancesco, R. (1995) J. Virol. 69, 1769–1777.
[26] Satoh, S., Tanji, Y., Hijikata, M., Kimura, K. and Shimotohno, K. (1995) J. Virol. 69, 4255–4260.
[27] Shimizu, Y., Yamaji, K., Masuho, Y., Yokota, T., Inoue, H., Sudo, K., Satho, S. and Shimotohno, K. (1996) J. Virol. 70, 127–132.
[28] Mori, A., Yamada, K., Kimura, J., Koide, T., Yuasa, S., Yamada, E. and Miyamura, T. (1996) FEBS Lett. 378, 37–42.
[29] Kakiuchi, N., Hijikata, M., Komoda, Y., Tanji, Y., Hirowatari, Y. and Shimotohno, K. (1995) Biochem. Biophys. Res. Commun. 210, 1059–1065.
[30] Laemmli, U.K. (1970) Nature 227, 680–685.
[31] Mach, H., Middaugh, C.R. and Lewis, R.V. (1992) Anal. Biochem. 200, 74–80.
[32] Steinkuhler, C., Tomei, L. and DeFrancesco, R. (1996) J. Biol. Chem. 271, 6367–6373.

Article No. jmbi.1999.2745 availab[illegible]nline at http://www.idealibrary.com on IDEAL® [illegible]Mol. Biol. (1999) 289, 371–384

JMB

The Solution Structure of the N-terminal Proteinase Domain of the Hepatitis C Virus (HCV) NS3 Protein Provides New Insights into its Activation and Catalytic Mechanism

Gaetano Barbato, Daniel O. Cicero, M. Chiara Nardi, Christian Steinkühler, Riccardo Cortese, Raffaele De Francesco and Renzo Bazzo*

IRBM "P. Angeletti"
Via Pontina km 30.600
00040 Pomezia, Roma, Italy

**Corresponding author*

The solution structure of the hepatitis C virus (BK strain) NS3 protein N-terminal domain (186 residues) has been solved by NMR spectroscopy. The protein is a serine protease with a chymotrypsin-type fold, and is involved in the maturation of the viral polyprotein. Despite the knowledge that its activity is enhanced by the action of a viral protein cofactor, NS4A, the mechanism of activation is not yet clear. The analysis of the folding in solution and the differences from the crystallographic structures allow the formulation of a model in which, in addition to the NS4A cofactor, the substrate plays an important role in the activation of the catalytic mechanism. A unique structural feature is the presence of a zinc-binding site exposed on the surface, subject to a slow conformational exchange process.

Keywords: HCV; NS3; serine proteinase; structure; NMR

Introduction

Hepatitis C virus (HCV) is recognised as the principal etiologic agent of parenterally transmitted non-A, non-B hepatitis (NANB-H: Choo *et al.*, 1989; Kuo *et al.*, 1989). The virus establishes a chronic infection that persists for decades in at least 85% of the infected individuals and up to 70% develop chronic active hepatitis. Chronic infection ultimately leads to the development of liver cirrhosis and hepatocellular carcinoma. Neither a vaccine against viral infection nor effective therapy for HCV associated chronic hepatitis has been developed to date. With an estimated world-wide population of infected people of more than 150 million, HCV represents one of the most widely spread and challenging viral infections to block.

The HCV virion has a positive strand RNA genome of about 9.6 kb that encodes a polyprotein of about 3000 amino acid residues (Houghton, 1996). The genetic organisation of HCV is similar to that of Flavi and Pestiviruses and it was classified as a separate genus of the Flaviviridae family. The analysis of the sequence of HCV reveals that it exists in at least six major genotypes and 11 subtypes (Simmonds, 1994). However, all known HCV polyprotein sequences share at least 71% identity. The structural protein "core" and the envelope glycoproteins E1 and E2 are released from the N-terminal portion of the polyprotein by action of cellular peptidases, while the non-structural proteins involved in the replication of HCV are released by the action of two virus-encoded proteinases: NS2-3 and NS3 (for reviews, see Bartenschlager, 1997; Neddermann *et al.*, 1997). NS2-3 is a zinc-dependent proteinase that performs a single proteolytic cut to release the N terminus of NS3. The proteolytic cleavage at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A, NS5A/NS5B junctions is mediated by a serine proteinase contained within the N-terminal 180 amino acid residues of NS3. The C-terminal (residues 180-630) of NS3 has been demonstrated to possess helicase activity. It has also been shown that the action of a cofactor, NS4A, enhances the activity of the serine proteinase in all the cleavages, *via* the formation of an

Abbreviations used: HCV, hepatitis C virus; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; SCR, structurally conserved regions.

E-mail address of the corresponding author: bazzo@irbm.it

0022-2836/99/220371–14 $30.00/0

NS3/NS4A complex. Interaction with the NS4A cofactor is required to perform the cleavages at NS3/NS4A, NS4A/NS4B and NS4B/NS5A junctions but the proteinase in its uncomplexed state is still able to cleave at the NS5A/NS5B boundaries, although with a much lower activity.

Since the NS3 proteinase is involved in the maturation process of the virus, the study of the structure of this enzyme is of crucial importance from a pharmacological point of view, in that it can give a strong impulse to the design of inhibitors that may prevent its action and thus block the viral replication and spread.

The crystallographic structures of the free enzyme (Love *et al.*, 1996; for simplicity in the following we refer to it as ns3) and its complex with a peptide representing the central region of the NS4A protein cofactor complex (Kim *et al.*, 1996; Yan *et al.*, 1998; we refer to it as ns3-4a) have been solved. The overall topology is similar in both structures, and forms an N-terminal (approximately residues 1-93) and a C-terminal (residues 94-180) six-stranded anti-parallel β-barrel. The barrels are packed like those of chymotrypsin-like serine proteinases. The catalytic site is formed by the triad of residues H57, D81 and S139, and is found in the crevice between the two domains. In addition to the β-barrels, there are two helical segments: α1 (residues 56-60), comprising the catalytic histidine residue, and α2 (residues 131-137) present also in the ns3 and ns3-4a structures. Two additional helices, α0 (residues 13-21) and α3 (residues 172-180), are formed only in the ns3-4a structure (Kim *et al.*, 1996; Yan *et al.*, 1998).

The commonly accepted mechanistic model of action of the serine proteinases implies a relay mechanism of hydrogen bonds involving, on one side, the carboxylate moiety of the Asp and the δ HN of the His residues and, on the other side, the ε N of the His and the γ HO of the Ser residues. This relay of H-bonds activates the γ O of the Ser residue, which can produce the nucleophilic attack on the C atom of the scissile bond (Fersht, 1984; Polgar, 1989; Lesk & Fordham, 1996).

In the ns3 structure, the carboxyl group of D81 is positioned far from and points away from H57 (Love *et al.*, 1996) impairing the H-bond formation. Conversely, in the ns3-4a structure, the side-chain of D81 is within hydrogen-bonding distance from the H57 imidazole group (Kim *et al.*, 1996; Yan *et al.*, 1998). By comparison of the ns3 and ns3-4a structures it has been inferred that D81 is correctly positioned as a member of the canonical catalytic triad as a consequence of the presence of the NS4A cofactor (Love *et al.*, 1998).

This conclusion, reasonable as it may seem from the crystallographic data alone, is challenged by new available biochemical evidence. In fact, the pH-dependence of the hydrolysis reaction studied in the presence of substrate and with or without NS4A titrates with the identical p*K* value of 7.2 (Landro *et al.*, 1997). The authors conclude that the activation role of NS4A is not exerted by a perturbation of the pK_a values of the active-site residues involved in the catalysis, in contrast with the model proposed by Love *et al.* (1998). On the other hand, NMR pH titration of the catalytic H57 residue of the free enzyme gives a pK_a value of 6.8 (Urbani *et al.*, 1998). Assuming that the pK measured by Landro *et al.* (1997) reflects the actual value for H57, the difference in pK_a value from that of the free enzyme could be due either to a direct role of the substrate itself in the pK_a alteration or to the different buffers used in the experiments. A kinetic analysis conducted on different types of substrate-like inhibitors in the absence and presence of the NS4A cofactor has shown that the action of NS4A peptide is exerted only on the P′-side of the substrate (Landro *et al.*, 1997). From these findings, the authors conclude that NS4A modulates NS3 activity by alteration of the S′ subsites.

At the moment, despite knowledge of the crystallographic structures in the absence and presence of the NS4A cofactor, its mechanism of activation and its role on the catalytic triad relative orientation is not completely understood. Here, we illustrate the first solution structure of the NS3 proteinase domain (first N-terminal 180 residues of NS3 with the addition of a solubilising six residue tail with the amino acid sequence ASKKKK) obtained in the absence of the NS4A cofactor and based on NMR data. The novelty of our findings is that the global architecture responsible for the relative positioning of the catalytic residues is already present in the absence of the NS4A cofactor. This difference from the relative crystallographic structure ns3 accounts for all the biochemical evidence available to date. The action of the cofactor is then discussed in terms of stabilisation of the fold of the N-terminal region and by its influence on the substrate leaving-group S′ side, while an influence on the substrate recognition S side can be excluded. Also, a possible role of the substrate in the relative positioning of the catalytic triad is proposed and discussed.

A rather unusual structural feature of the NS3 enzyme is the presence of a zinc-binding site completely exposed to the solvent. We find that this site in solution undergoes a conformational exchange between an open and a closed conformation by switching the side-chain of H149 on the hundreds of milliseconds timescale.

Results and Discussion

Structure determination

The solution structure of the protease domain of the hepatitis C virus NS3 protein of the strain BK was solved by multidimensional heteronuclear NMR spectroscopy (Clore & Gronenborn, 1991a; Bax & Grzesiek, 1993) making use of uniformly labelled ^{15}N, $^{15}N/^{13}C$ and $^{2}H/^{15}N/^{13}C$ samples, as well as of a selectively ^{15}N[Leu], ^{15}N[Val] and ^{15}N[Ala]-labelled sample. Complete resonance assignment was obtained, except for 15 residues at the N terminus and few other signals. In fact, the

proton resonances of residues 6-21 appear to be broadened beyond detection by conformational exchange. Leucine residues 14, 15 and 21 resonances, for example, were identified as very broad peaks in a sample selectively labeled with leucine but it was not possible to identify them sequentially. The general quality of NMR data is shown in Figure 1, which depicts strips from (a) a 3D ^{13}C edited NOESY and (b) from a ^{15}N edited NOESY. The structure was calculated, excluding the first 21 residues, by simulated annealing (Nilges *et al.*, 1988) using the database of constraints shown in Table 1, where a summary of the structural statistics is given. In Figure 2, a stereoview of the overlay of the 20 lower-energy structures generated is shown. For simplicity we will refer in the following to the minimised average solution structure as nmr.

Description of the structure

The nomenclature used for trypsin β-strands and applied to the topology of the protein is shown in Figure 3(a). The sequence alignment of NS3 with several other chymotrypsin-like serine proteinases is proposed in Figure 3(b), on the basis of the structural overlay, following the guidelines indicated by Greer (1990). This alignment differs slightly from that proposed previously (Love *et al.*, 1996). The

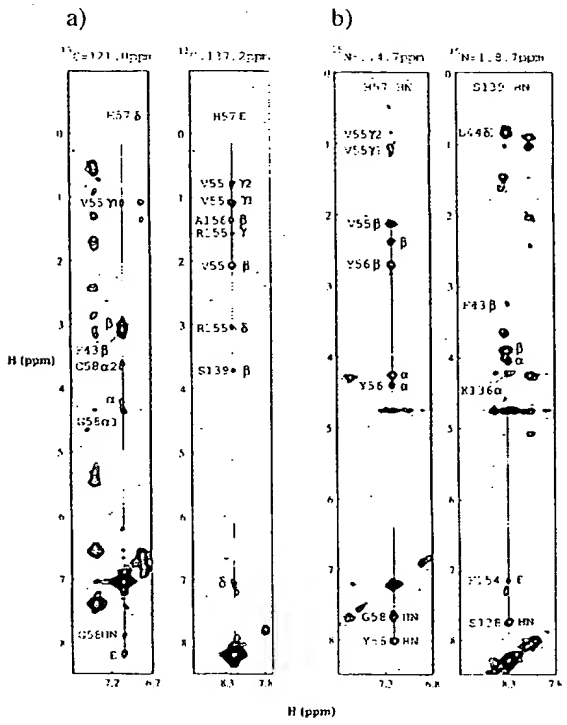


Figure 1. Examples of spectra showing the quality achieved with this protein. (a) Strips taken from a ^{13}C edited 3D NOESY show the aromatic H^{ε} and H^{δ} NOEs of the catalytic histidine residue. (b) Strips of the amide protons of both the catalytic histidine and serine residues from a ^{15}N edited 3D NOESY.

Table 1. Experimental restraints and structural statistics

A. *NMR constraints*		
NOE	Total	2476
	Intra	945
	Inter short distance ($<i+3$)	521
	Inter long-range ($>i+3$)	1010
Generic	Total	70
	H-bond	64
	Zn-binding site	6
Dihedral	Total	83
	ϕ	37
	$\chi 1$	43
Stereospecific	Methylene groups	31/219
	Methyl groups	50/66
B. *Structure statistics*		
R.m.s. deviations from experimental constraints[a]		
Distance (Å)		0.076 ± 0.003
Dihedral(deg.)		1.331 ± 0.148
$^{13}C^{\alpha}$		1.49 ± 0.05
$^{13}C^{\beta}$		1.09 ± 0.06
Deviations from idealized geometry		
Bonds (Å)		0.005 ± 0.0006
Angles (deg.)		0.761 ± 0.032
Impropers (deg.)		0.543 ± 0.01
Coordinates precision referred to mean structure (Å)		
Residues SCR+helices		
Backbone		0.472 ± 0.089
All heavy atoms		1.147 ± 0.160
All residues		
Backbone		0.872 ± 0.097
All heavy atoms		1.306 ± 0.117
N-terminal residues SCR+helices		
Backbone		0.554 ± 0.140
All heavy atoms		1.020 ± 0.292
C-terminal residues SCR+helices		
Backbone		0.233 ± 0.042
All heavy atoms		0.567 ± 0.091
C. *Ramachandran analysis*[b]		
% Residues in most favoured regions		70.4 ± 1.8
% Residues in allowed regions		26.7 ± 1.4
% Residues in generously allowed regions		3.0 ± 0.6
% Residues in disallowed regions		0.0 ± 0.0

[a] None of the structures exhibited distance violations greater than 0.5 Å or dihedral angle violations greater than 5°.

[b] The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the overall quality of the structures. The residues with a heteronuclear NOE ^{15}N-$^{1}H < 0.6$ (total 28 residues) were excluded from the computations because of their intrinsic mobility.

structurally conserved regions (SCR) are indicated with boxes in the alignment. The overall sequence similarity is very low (<20%), but the general topology is well conserved. The NS3 proteinase, like the proteinase from Sindbis virus (Tong *et al.*, 1993) and from the Semliki Forest virus (Choi *et al.*, 1997), is a small proteinase (about 180 residues) and, as such, makes an economical use of loops, lacking all of a series of connecting elongations that are a common feature of cellular proteinases: we will see below that this has a peculiarly relevant consequence.

The N-terminal β-barrel appears to be less compact than the C-terminal one (Figure 2). Evidence for this is obtained from a comparison of the num-

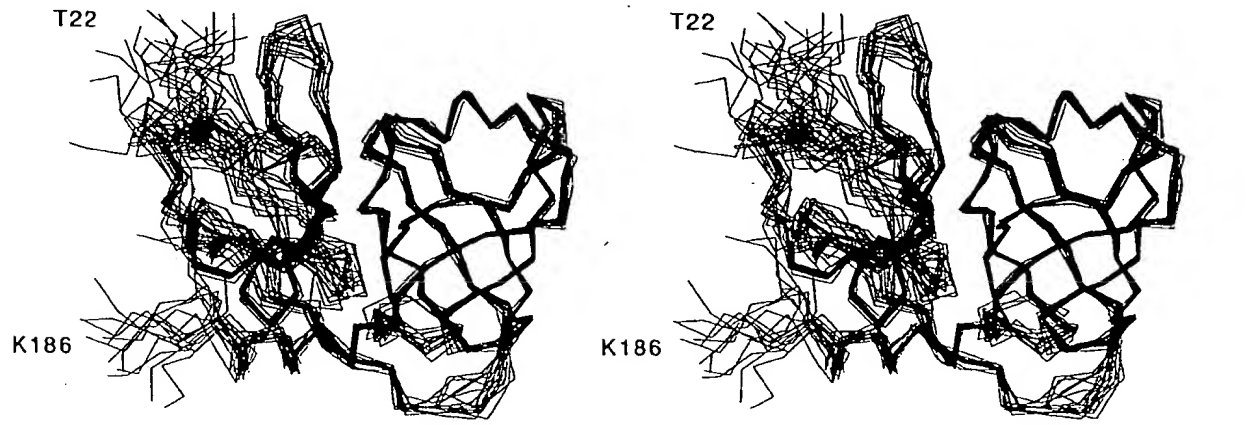


Figure 2. A stereoview of the 20 minimum-energy structures is shown. For the overlay, the SCR residues identifying the β-strands plus the helicoidal segments were used. Due to its peculiar mobility, see the NS4a interaction section, strand D1 was omitted in the calculation of the r.m.s.d. The first 21 residues were not included in the structure calculation, because for them no structural information was available.

ber of slow-exchanging amide protons between the two domains: 12 in the N-terminal and 20 in the C-terminal domain. In Figure 4(a), (b) and (c) the number of observed NOEs per residue, the backbone heavy atoms r.m.s.d. per residue, and the heteronuclear ^{1}H-^{15}N NOEs are reported, respectively. The total number of NOEs for the N-terminal barrel is 990, whereas for the C-terminal it is 1486. The ^{1}H-^{15}N NOE values in 17 residues in the N-terminal portion of the molecule are below 0.6 (indicating a high level of mobility), while this is the case for only six residues in the C-terminal domain (excluding the residues forming the solubilising tail). These differences are reflected in the poorer r.m.s.d. of the N-terminal barrel compared to the C-terminal one (Figure 4(b)).

Table 2 illustrates the pairwise comparison of the SCR residues between the nmr and the ns3 and ns3-4a structures. The r.m.s.d. values for the backbone heavy atoms are 1.45 and 1.18 Å, respectively. However, if we consider the N and C-terminal β-barrels separately, we get a better insight into the comparison. There is a substantial difference between the N-terminal β-barrel, with an r.m.s.d. of 1.68 Å (nmr/ns3), of 1.48 Å (nmr/ns3-4a) and of 1.98 Å (ns3/ns3-4a). The nmr structure looks different to a similar extent from either of the two crystallographic structures, thereby suggesting that, in the crystal of the uncomplexed enzyme, crystal forces are responsible for distorting the N-terminal β-barrel. The C-terminal β-barrel with an r.m.s.d. of 0.52 Å (nmr-ns3), of 0.56 Å (nmr/ns3-4a) and of 0.41 Å (ns3/ns3-4a) is very similar in all three structures. However, a major difference in this domain is given by the fact that in solution (thus in the absence of the NS4A peptide) we find helix α3 (residues 172-182), while this helix is absent from the ns3 (in the absence of NS4A peptide) structure. In Figure 5(a) a zoomed view of the overlay of the ns3 and nmr structures is shown, providing evidence for this difference, which is extremely relevant in the evaluation of the role played by NS4A peptide to activate the proteinase and will be further discussed in the catalytic triad and substrate-binding section.

Table 2 also reports the pairwise comparisons with several other serine proteinases belonging to the chymotrypsin family. Also in this case it is instructive to consider separately the two β-barrels. The fold of the N-terminal SCR residues in solution is similar to that of chymotrypsin (Blevins & Tulinsky, 1985), trypsin (Bode *et al.*, 1984) and elastase (Meyer *et al.*, 1988), with an r.m.s.d. of 1.37, 1.40 and 1.42 Å, respectively, while the ns3 structure gives 1.80, 1.91 and 1.84 Å, respectively, and the ns3-4a structure gives 2.12, 2.12 and 2.11 Å, respectively. One can conclude that, in absence of the NS4A peptide, the overall fold of the β-barrel is conserved; when forming the complex with the NS4A peptide, the N-terminal β-barrel undergoes a substantial change in conformation (specifically, the D1 strand and the loops preceding and following it; Figure 3(a)) that differentiates it locally from the other proteinases belonging to the same family.

The C-terminal β-barrel of all the NS3 structures, although very similar to each other, differs from those of the other chymotrypsin-like proteinases (r.m.s.d. >2.0 Å). This is partially due to the slightly different packing of the E2 strand, which is directly involved in substrate binding, and of the C2 strand, which is packing directly against it (Figure 3(a)). This difference can be accounted for by the peculiarity of the substrate recognition surface for NS3. In fact, substrate specificity studies have shown that the NS3 proteinase requires at least a decamer peptide spanning P6-P4′ for optimal activity. The substrate frame needed thus is unusually long for serine proteinases.

Incidentally, the only other viral serine proteinases, Sindbis virus (Tong *et al.*, 1993) and Semliki Forest virus core protein (Choi *et al.*, 1997), for which a structure is available in the Brookhaven database, are more similar to NS3 with an r.m.s.d. for the C-terminal β-barrel

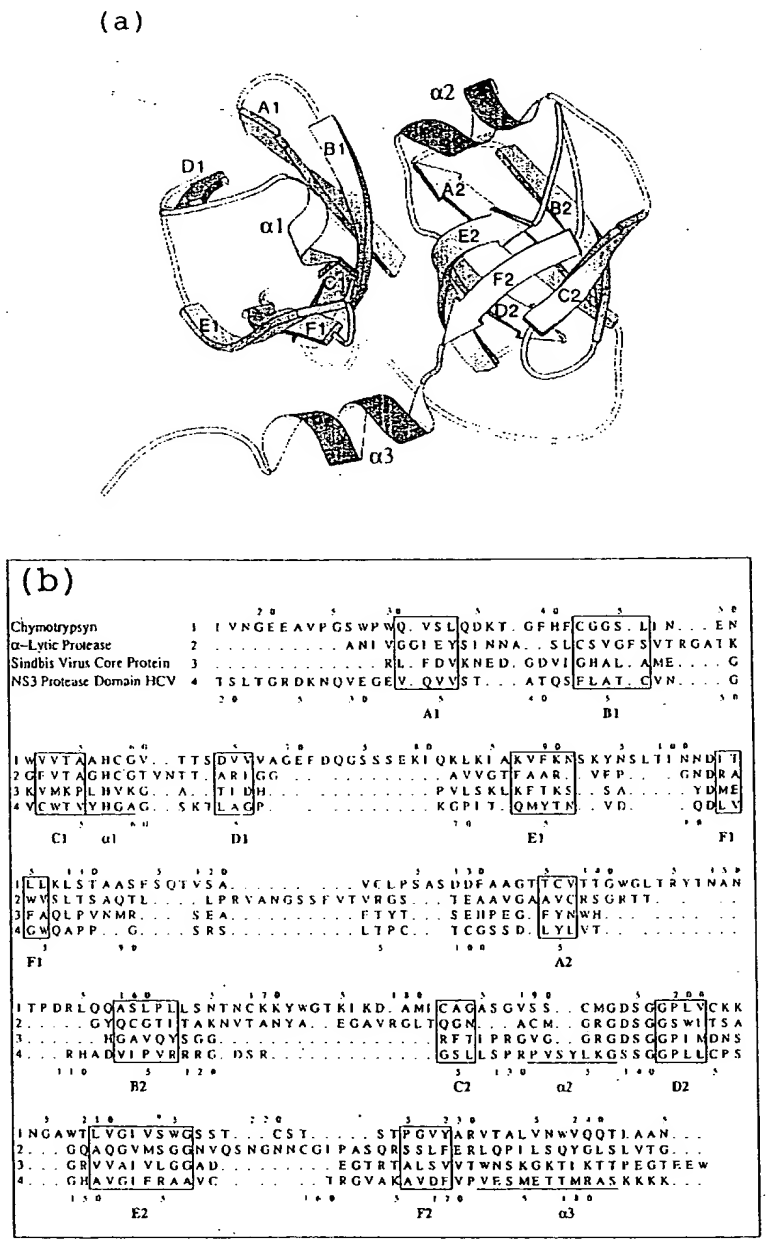


Figure 3. Topology and sequence alignment. (a) The topology of the NS3 proteinase domain is represented in a MOLSCRIPT view. The SC Regions are constituted by the strands A1-F1 in the N-terminal and A2-F2 in the C-terminal β-barrel. The helices are named α1, α2 and α3. (b) The structure alignment of NS3 with chymotrypsin-like serine proteinase. The alignment results from the best superposition of the C^{α} of the SCR residues, in the box. The upper numeration is referred to the chymotrypsin, while the bottom is referred to the NS3 proteinase domain. In bold are the residues of the catalytic triad, while the asterisk (*) at the top indicates the position of residue S214, which is strictly conserved in all cellular proteinases, and its corresponding residue in the two viral proteinases. Below the NS3 amino acid sequence is shown the corresponding secondary structure elements; strands are coincident with the boxed regions, while the helices are underlined.

ranging between 1.6 and 1.72 Å for all the pairwise comparisons.

The positioning of NS4A

The lower definition of the N-terminal β-barrel in the solution structure should be related to the absence of the NS4A protein cofactor. In Figure 6, a zoomed view of the overlay between the nmr (blue) and the ns3-4a (NS3 cyan, NS4A magenta) structures is shown. It has been shown by deletion mutagenesis experiments (Failla *et al.*, 1995), and more recently by the ns3-4a crystallographic structure (Kim *et al.*, 1996; Yan *et al.*, 1998), Figure 6, that only the N-terminal β-barrel of the proteinase is involved in binding the NS4A peptide cofactor. According to these structures, the NS4A peptide cofactor is almost completely buried inside the core of the N-terminal β-barrel, where it forms a β-strand within a four-stranded β-sheet. Its companion strands are formed by residues 4-10 (strand A0) and 33-37 (A1). Residues 13-21 form a short α-helix (α0) that is likely to contribute to the stability of the NS3-NS4A complex in the crystal. This helix is very peculiar, since the residues exposed to the solvent are three leucine and two isoleucine residues (Figure 6). It is unlikely that such a structure could exist in solution. In this respect, it is interesting to note that only one of the monomers in the crystallographic asymmetric unit is folded in the way just described (Kim *et al.*, 1996). In the second monomer, the first 30 residues do not have a

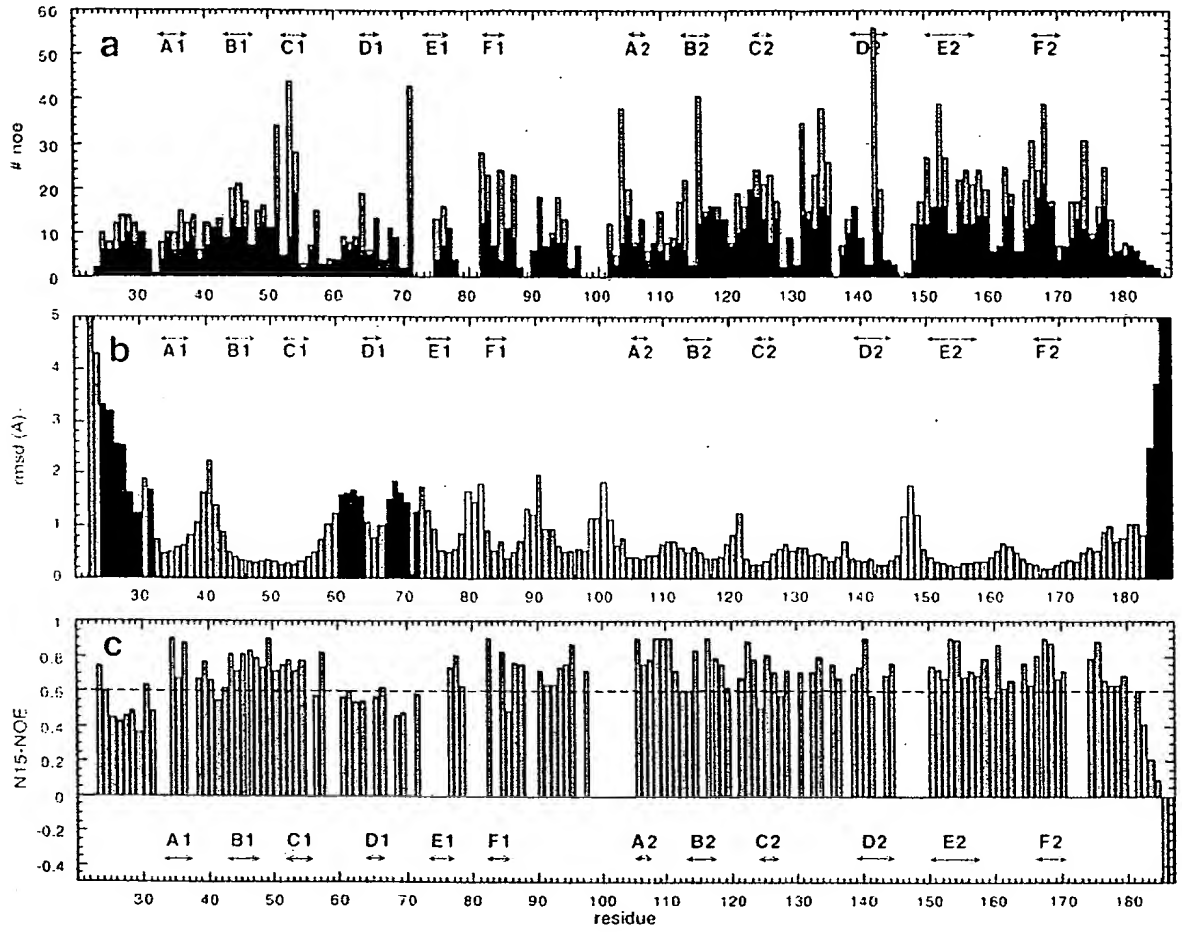


Figure 4. Correlation between the number of NOE constraints, mobility and r.m.s.d. The SCR strands are defined by the arrows and reported in all three panels to simplify reading the Figure. (a) A representation of the number of NOEs per residue, in black are represented the NOEs involving backbone protons, while in grey are represented those involving the side-chains proton. (b) Behaviour of the r.m.s.d. per residue. Regions in which the r.m.s.d. is poor can be related to low values of the ^{1}H-^{15}N NOE (<0.6) and are in black, or to a lack or low number (<10) of NOE constraints, and are in white. Only two regions do not follow this characteristic, namely residues 39-41 and 120-122, both are well-characterized turns but lack long-range NOEs. (c) Heteronuclear ^{1}H-^{15}N NOE value per residue number. The broken line drawn at the value of 0.6 indicates the mobile regions (<0.6).

defined structure. Despite these differences, both complexes contain the NS4A peptide in essentially the same position. It could be argued that more than one conformation can be assumed by the N-terminal 30 amino acid residues of NS3 even in the presence of NS4A. In our experiments, residues 6-21 signals are broadened almost beyond detection, moreover residues 25-31 exhibit a high mobility in solution (Figure 4(c)). Therefore, it seems reasonable to assume that some type of mobility is affecting all the N-terminal 31 residues. Evidence obtained by limited proteolysis experiments suggest indeed that the N-terminal region of NS3 is highly accessible in solution, even in the complex with the NS4A peptide cofactor (data not shown).

The N-terminal residues of the NS4A peptide contact directly the D1 strand and the preceding and following loops α1-D1 and D1-E1, respectively (Figure 6). In solution, these regions together with the strand D1 itself are characterised by a high degree of mobility (^{1}H-^{15}N NOE < 0.6; Figure 4(c)). From the overlay of the nmr and ns3-4a structures shown in Figure 6, it is evident that the whole region encompassing the residues 61-66 (loop α1-D1 and strand D1) are held down toward the F1 strand in the crystallographic structure, while in solution this strand is packing across the A1 strand, compensating for the absence of the NS4A peptide. From this comparison it can be concluded that one of the roles exerted by NS4A is to stabilize the strands A1, D1, the loop α1-D1 and D1-E1 in a more defined conformation thus compacting the whole N-terminal barrel. Its influence on the substrate-binding region is primarily due to the direct interaction and consequent conformational stabilisation of the strand A1 and the region α1-D1, which form the walls surrounding the P′-side of the substrate (see the next section).

Table 2. The r.m.s.d. comparison with several serine proteinases

N-terminal									
	2snv	5cha	2alp	1ntp	3est	2sga	3sgb	ns3	ns34a
nmr	1.91	1.37	1.86	1.40	1.42	1.70	1.82	1.68	1.48
2snv		1.45	2.12	1.60	1.50	2.25	2.03	1.77	2.45
5cha			1.81	0.36	0.25	1.67	1.73	1.80	2.12
2alp				1.78	1.78	1.09	0.37	2.02	2.17
1ntp					0.33	1.64	1.71	1.91	2.12
3est						1.66	1.70	1.84	2.12
2sga							1.07	1.96	1.77
3sgb								1.90	2.11
ns3									1.98
C-terminal									
	2snv	5cha	2alp	1ntp	3est	2sga	3sgb	ns3	ns34a
nmr	1.72	2.20	1.86	2.18	2.25	2.14	2.13	0.52	0.56
2snv		1.51	1.66	1.51	1.52	1.61	1.61	1.60	1.66
5cha			0.75	0.45	0.49	0.72	0.79	2.13	2.18
2alp				0.76	0.80	0.42	0.47	2.06	2.14
1ntp					0.63	0.69	0.73	2.11	2.18
3est						0.78	0.83	2.16	2.22
2sga							0.24	2.02	2.10
3sgb								2.02	2.10
ns3									0.41
all									
	2snv	5cha	2alp	1ntp	3est	2sga	3sgb	ns3	ns34a
nmr	2.09	2.08	2.15	2.03	2.09	2.06	2.11	1.45	1.18
2snv		1.66	2.06	1.74	1.63	2.07	1.99	1.95	2.27
5cha			1.51	0.46	0.52	1.40	1.44	2.04	2.31
2alp				1.44	1.47	0.83	0.47	2.14	2.25
1ntp					0.60	1.33	1.38	2.06	2.26
3est						1.40	1.44	2.08	2.31
2sga							0.77	2.12	2.03
3sgb								2.04	2.21
ns3									1.66

Root-mean-square deviation (Å) comparison of the SRC backbone residues heavy atoms for several serine proteinases. The names are given as Brookhaven PDB codes: 1bt7 (nmr); 1a1q (ns3); 1jxp (ns3-4a); 2snv (Sindbis virus core protein); 5cha (bovine alpha-chymotrypsin); 2alp (alpha lytic protease); 1ntp (bovine beta trypsin); 3est (porcine elastase); 2sga (*Streptomyces griseus* protease A); 3sgb (*Streptomyces griseus* protease B). Due to the high degree of similarity between the Sindbis virus and the Semliki Forest virus core proteins (1vcp) (r.m.s.d. 0.4 Å), the comparison are reported only for the former.

The high level of similarity between the C-terminal barrel of the solution and that of the crystallographic structures, is quantified by the low r.m.s.d. of the backbone heavy atoms for the SCR residues (Table 2). It can be concluded that the complex with the NS4A peptide has little, if any, influence on this region of the enzyme, which contains the recognition pockets for the P-side residues of the substrate.

These conclusions find experimental support from a steady-state kinetic analysis of inhibitor binding to the active site of the NS3 proteinase (Landro *et al.*, 1997). In fact, two classes of competitive inhibitors could be identified: those interacting

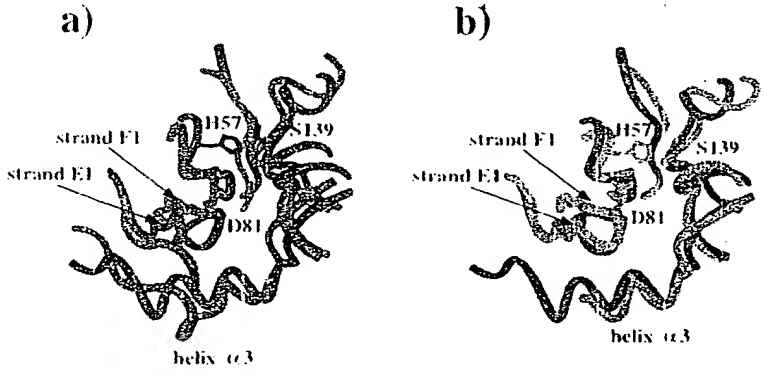


Figure 5. Pairwise superposition of the nmr (blue) with ns3 (red) and ns3-4a (cyan) structures. The superposition is obtained from the r.m.s.d. of all the SCR residues. The catalytic triad residues are shown on only one of the structures to facilitate the comparison. The position of D81 in the nmr structure is indicated by a dot, since its position is not accurately determined. The residues were not shown for ns3, since only the C° coordinates are publicly available. (a) nmr/ns3: the strands E1-F1 containing the loop that bears the catalytic Asp are different. In the ns3 structure, the helix α3 is essentially absent. (b) nmr/ns3-4a: the strands E1-F1 bearing the Asp residue are similarly positioned, and the helix α3, which is packed mainly against the strand E1, is very similar in both structures.

only with the P binding pockets located on the C-terminal β-barrel and those compounds extending their interaction with the enzyme to both the P and P′ binding sites. The potency of the former category of inhibitors was not influenced by complex formation with NS4A. In contrast, the affinity of active-site ligands relying only on contacts with the P′ binding site on the N-terminal β-barrel was strongly impaired in the absence of NS4A.

The catalytic triad and substrate-binding region

The serine proteinases of the chymotrypsin family share a number of elements: a catalytic triad formed by residues Ser-His-Asp; a site for hydrogen bonding to a tetrahedral oxyanionic intermediate of the reaction (also called the oxyanion hole); a strand forming an antiparallel β sheet with the P-side of the polypeptide chain of the substrate (also called S site), which contributes also to the formation of a recognition pocket, and a leaving-group recognition site (also called S′ site). Each of these elements occurs in the different members of the family in an almost identical geometric relationship.

The alignment with other serine proteinases shows that these residues and the nearby positions are well conserved (Figure 3(b)). From the structural point of view, the relative position of H57 and S139 is similar in the nmr and ns3-4a structures and only slightly more apart than in the other serine proteinases (Table 3). On the contrary, in the ns3 structure "the imidazole of H57 [...] is oriented toward S139 but is not close enough to form the H-bond observed in proteinase structures" (Love *et al.*, 1996). The difference in distance (Table 3) is very likely due to distortions induced by the crystal forces, since in solution we also find that the observed pK_a value of 6.8 for H57 (Urbani *et al.*, 1998) (uncomplexed enzyme) is in agreement with the catalytic histidine residue being H-bonded with the catalytic serine, as expected from the canonical model of serine proteinases.

In the nmr structure, the position of D81 is not accurately determined because of the lack of experimental constraints. As a matter of fact, for the resonances of residues 79, 80 and 81 no inter-residue

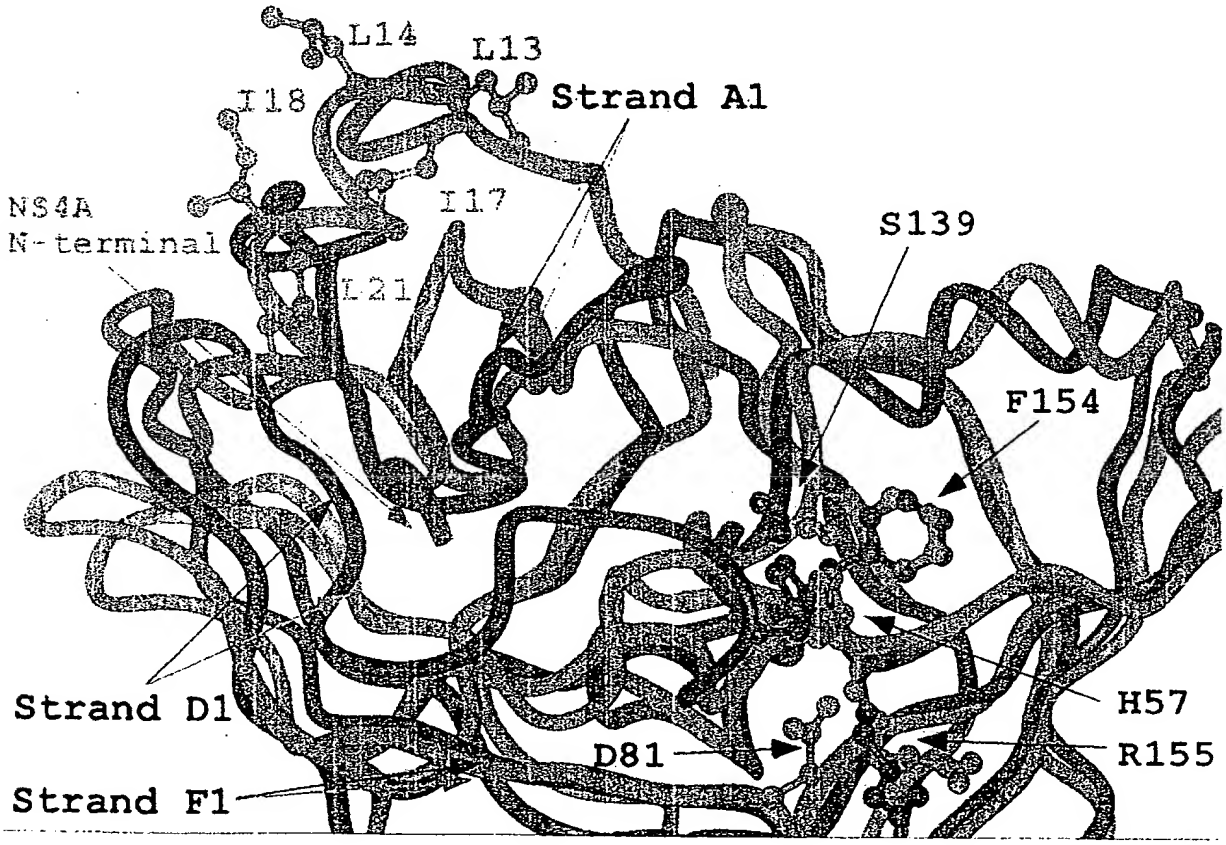


Figure 6. The role of NS4A; zoomed vision of the superposition of nmr/ns3-4a structures (blue/cyan-magenta). The catalytic triad residues as well as R155 side-chains are shown to facilitate the comparison. The role of NS4A seem to be to order the N-terminal 21 residues in an initial strand and a subsequent helix that turns around the N terminus of NS4A. However, this helix exposes hydrophobic residues to the solvent while packing inside the hydrophilic residues. A situation like this is not possible in solution. The positioning of NS4A has the consequence of order-

Table 3. Geometric relation of the catalytic residues in serine proteinases

	His-Ser (Å)	His-Asp (Å)	Asp-Ser (Å)
nmr	9.0	-	-
ns3	9.7	7.9	11.5
ns3-4a	9.3	6.5	10.5
2snv	8.8	7.0	10.6
5cha	8.5	6.5	10.5
1ntp	8.5	6.3	10.1
2alp	8.3	6.2	9.9
3est	8.4	6.5	9.8
2sga	8.5	6.2	9.9
3sgb	8.4	6.3	10.1

Comparison of the distances between C^{α} of the catalytic triad in NS3 structures and in several serine proteinases. The actual distances His-Asp and Asp-Ser for the nmr structures are not included, since the whole loop containing D81 undergoes substantial conformational averaging, so that a single evaluation of distance would not be meaningful.

NOE is observed and the signals arising from the amide protons of these residues are affected by their fast water exchange rates. The resulting calculated structures show that several allowed conformations of this loop are sampled in the simulated annealing computations, all of them solvent-exposed.

It is worthwhile to point out, that also in the ns3-4a crystal structure this loop is solvent-exposed and the temperature factors of the backbone atoms of the three residues 79-81 are around 50.0 Å^2, which indicates a high degree of mobility even in presence of the NS4A peptide. Moreover we find that, in the average solution structure of the isolated NS3 proteinase domain, the strands E1 and F1, which enclose the loop bearing D81, are positioned similarly to the crystal ns3-4a structure despite the absence of the cofactor, as clearly shown in the zoomed view of the overlay between the nmr and ns3-4a structures in Figure 5(b). These strands are held in position by the packing of the helix α3 (residues 172-182), which in solution is similar to that in the ns3-4a crystal (Figure 5(b)). On the contrary, in the ns3 structure this helix is absent (Figure 5(a)), strongly suggesting that the crystal packing is inducing a distorted conformation in this region. The consequence is that the strands E1 and F1 that position the D81 are distorted and thus its positioning in the ns3 structure is not compatible with a well-formed catalytic triad. Our results in solution thus are in contrast with the conclusions drawn by Love *et al.* (1998) based on the crystallographic structures alone. On the other hand, our results are in agreement with the biochemical evidence that the presence of NS4A is not affecting the pK_a of the catalytic residues (Landro *et al.*, 1997), indicating that the catalytic triads of the enzyme-substrate complex and of the ternary enzyme-substrate-NS4A complex must possess very similar geometries.

From all the above considerations, we conclude that other factors in addition to the presence of NS4A are to be invoked to stabilize the position of the D81 residue in a canonical catalytic triad configuration. The identification of such factors could be attempted by analysing the structures of several members of the chymotrypsin-like family of serine proteinases. Ideally, we can divide it into two subfamilies; namely, the short-chain, about 180 residues, and the long-chain proteinases, about 250 residues (Bazan & Fletterick, 1998). In all the solved long-chain structures there are two conserved characteristics: (i) the residue at position 214 is invariably serine, which forms an H-bond with the carboxylic group of the catalytic aspartate residue, thus helping in its correct positioning and alignment in respect to the catalytic histidine; (ii) the catalytic aspartate residue is shielded from the solvent by the hydrophobic residues at conserved positions. In Table 4 these characteristics are summarized for several long-chain serine proteinases and the short-chain proteinases for which a structure is available. They do not share these characteristics, and incidentally these proteinases are all of viral origin.

The protection from the solvent allows the direct observation of the δ HN proton NMR signal of the

Table 4. Summary of some serine proteinases that present both Ser in position 214 and the catalytic Asp sheltering from the solvent, and the few known exceptions

Serine proteinase	PDB entry	Category	Origin	214	Sheltering	Reference
Trypsin	1NTP	LC[a]	Mammalian	Ser	Y94, L99	Bode *et al.* (1984)
Chymotrypsin	5CHA	LC	Mammalian	Ser	Y94, I99	Blevins & Tulinsky (1985)
Kallikrein	2PKA	LC	Mammalian	Ser	F94, Y99	Bode *et al.* (1983)
Thrombin	1HXE	LC	Mammalian	Ser	Y94, L99	Bode *et al.* (1989)
Mast cell proeinase	3RP2	LC	Mammalian	Ser	Y94, N99	Remington *et al.* (1988)
Tonin	1TON	LC	Mammalian	Ser	Y94, L99	Fujinaga & James (1987)
Elastase	1HNE	LC	Mammalian	Ser	Y94, L99	Meyer *et al.* (1988)
Streptomyces griseus A	2SGA	LC	Bacterial	Ser	F94, Y171, V177	Moult *et al* (1985)
Streptomyces griseus B	3SGB	LC	Bacterial	Ser	F94, Y171, V177	Read *et al.* (1983)
Alpha lytitc protease	2ALP	LC	Bacterial	Ser	F94, Y171, V177	Fujinaga *et al.* (1985)
Sindbis core protein	2SNV	SC[a]	Viral	Leu		Tong *et al.* (1993)
Semliki Forest core protein	1VCP	SC	Viral	Leu		Choi *et al.* (1997)
NS3(1-180) HCV	1A1R 1JXP	SC	Viral	Arg		Kim *et al.* (1996) Yan *et al.* (1998)

[a] LC, long-chain, SC, short-chain Serine proteinases.

histidine residue involved in the H-bond with the carboxylic moiety of the aspartate residue (Frey *et al.*, 1994). This proton resonates at an unusually low field chemical shift, in the range of 14.5-19.0 ppm depending on pH, and has been observed in several serine proteinases of the long-chain family (Markley, 1978; Bachovchin, 1985; Frey *et al.*, 1994). In the case of the NS3 proteinase, however, despite all attempts, such a signal has until now not been observed. From both the nmr and ns3-4a structures, one could argue that this is due to the site being solvent accessible. We speculate that, in solution, the aspartate residue is unlikely to be engaged in an H-bond with the histidine residue, even in the presence of the NS4A cofactor. However, an environment similar to that of the long-chain subfamily could be partially created with the participation of the substrate. We observe, in fact, for the short-chain enzymes a tendency to have a hydrophobic bulky residue in position P2 (Tong *et al.*, 1993; Ingallinella *et al.*, 1998), whereas long-chain enzymes tend to have either glycine or short side-chain residues such as Ala (Yasutake & Powers, 1981; Chang, 1986; Coombs *et al.*, 1996). A bulky residue in position P2 together with the aliphatic methylene groups of R155 (NS3) and of the L231 side-chain (Sindbis and Semliki viruses) could contribute to shelter the aspartic acid side-chain, thereby favouring the formation of a catalytic triad machinery that more closely resembles that observed in long-chain enzymes.

The pH titration data obtained in the presence of substrate (Landro *et al.*, 1997) show that the value of pK_a 7.2 of the catalytic residues of the NS3 proteinase is NS4A-independent, but this value is different from that of 6.8 found for the catalytic histidine residue of the free enzyme (Urbani *et al.*, 1998). One could speculate thus that a role in the enhancement of the pK_a value is played by the substrate itself; however, it is possible that the difference observed is due to the different buffer conditions used in the experiments. Preliminary spectroscopic evidence in agreement with the hypothesis of an involvement of the substrate itself in the conformational stabilization of the catalytic triad has been collected on substrate-based inhibitors (Cicero *et al.*, 1999).

The residues preceding the catalytic serine residue form the oxyanion-stabilizing loop (residues 137-139; Figure 7). The S site comprises strand E2, which forms one side of the specificity pocket, with the A157 amide proton and carbonyl group accessible to form backbone to backbone H-bonds with the substrate P3 residue according to a classical chymotrypsin-like substrate interaction. The recognition pocket is shallow and apolar, being formed by the methyl groups of A157 and L135 on the side and by the F154 aromatic ring at the bottom (Figure 7). Features of this pocket have been predicted from modelling studies (Pizzi *et al.*, 1994).

The S′ site is constituted by the ending of the A1 (T38) and the beginning of the B1 strands, as well as the A1-B1 short loop (Figure 7). Strand A1 packs directly with the NS4A-activating peptide in the ns3-4a structure, thus explaining a direct influence of the NS4A peptide on the P′ portion of the substrate (leaving group). A more detailed insight into the S binding site, the detailed identification of the residues involved and their role in the interaction with substrate-based inhibitors are presented in the accompanying paper (Cicero *et al.*, 1999).

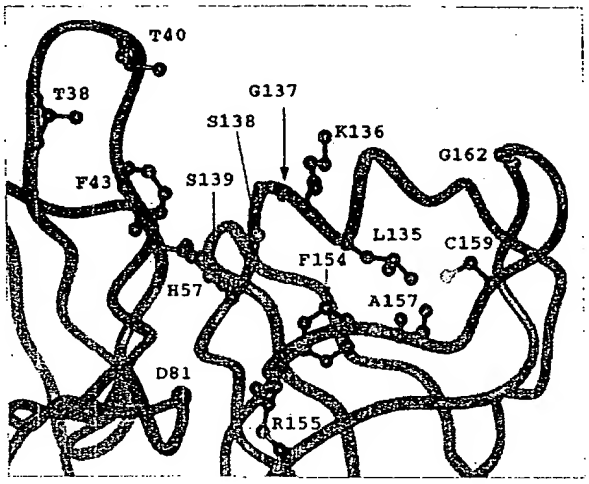


Figure 7. Zoomed view of the substrate interaction region. The S′ and S regions range approximately from strand A1 (identified by the position of T38) to loop E2-F2 (G162), encompassing the rather flat surface defined by the E2 strand (F154 bottom of recognition pocket, A157 H-bond candidate with substrate P3 partner) on one side and the α2 helix (oxyanion hole G137-S139; L135 delimiting the top side of the recognition pocket) on the other side.

The Zn-binding site

In a previous study, we showed that a zinc ion is required for the structural integrity and activity of the NS3 proteinase and, from modelling studies, the coordination site was predicted to be comprised of loops F1-A2 (C97 and C99) and B2-C2 (C145 and H149: De Francesco *et al.*, 1996). Subsequent publication of the crystallographic structures confirmed the predictions. In the crystallographic structures, the zinc coordination is essentially tetrahedral. However, the imidazole moiety and the zinc atom are too distant to be directly bound, and the authors postulated the presence of a water molecule acting as a bridge (Love & Hostomska, 1996; Kim *et al.*, 1996).

The zinc-binding site in solution, as determined by NMR, is illustrated by Figure 8. Our data identify the N^{δ} of H149 as the one involved in the binding to the zinc ion. This structural result is in agreement with our previous findings on the tautomeric state of H149, with N^{ϵ} being in the α state and N^{δ} being in the β state (Urbani *et al.*, 1998). We

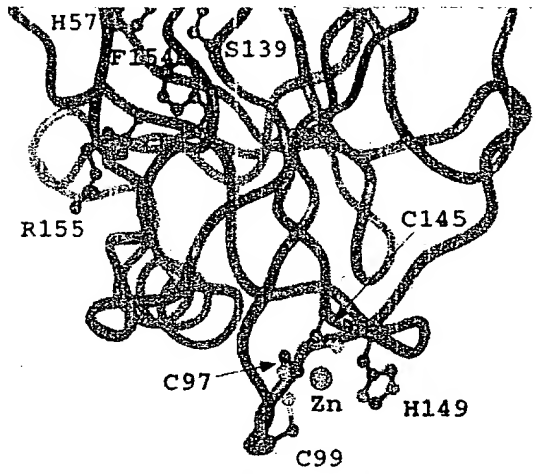


Figure 8. Zoomed view of the zinc-binding site. The residues chelating the zinc ion are shown; i.e., it is evident the unusually long stretch that links C145 to H149. This longer linker could be required to allow the conformational switch of H149 and leave space accessible to the zinc ion.

also found that the imidazole moiety of H149 modulates the accessibility of the zinc ion, allowing an "open" and a "closed" conformation in the protonated and unprotonated state, respectively, which interconvert on the 100 ms timescale (Urbani *et al.*, 1998). In this respect, one should notice that in the ns3 structure (Love *et al.*, 1996) H149 is postulated to participate in metal coordination in only two of the three molecules in the asymmetric unit, whereas in the third one the imidazoyl side-chain moves away. In the experimental conditions used in this structural study, the closed conformation is dominant. In the NMR-derived structure, the imidazole moiety was not subject to specific restraints that would force the coordination with the zinc atom (see Materials and Methods) and in all the resulting structures it is positioned at a distance too far to directly chelate the zinc atom. This result is in agreement with our previous findings, in which H149 is ligated to the metal using the δ1 N through a hydroxyl group (Urbani *et al.*, 1998).

Most of the chymotrypsin-like proteinases have disulphide bridges that are believed to maintain the relative orientations of the residues involved in catalysis (Lesk & Fordham, 1996). Disulphide bridges present in these extracellular serine proteinases are unlikely to be stable in the reducing intracellular milieu. Since NS3 is an intracellular proteinase, we proposed that the zinc-binding site is used to stabilise the relative orientation of the two β-barrel domains, thus indirectly influencing the position of the catalytic triad, which is located in the crevice between the domains (De Francesco *et al.*, 1996).

However, we observe also that the histidine residue is highly conserved in all the HCV strains and HCV-related viruses. This fact, together with the unusual location of the site, i.e. completely solvent-exposed, and its dynamic behaviour in solution may suggest that there could be a biologically relevant function, not yet clarified, linked to this zinc-binding site. Hepatitis C is a small virus that encodes only six non-structural proteins. Therefore, as observed in other viruses, multi-functional proteins are a strategy to minimise the number of agents needed during viral replication. Then the optimisation of the role of the zinc ion and the peculiar features of its binding site would be just part of the same biomolecular strategy.

Materials and Methods

Expression, purification and solubilisation

Escherichia coli cells BL21(DE3) were transformed with a plasmid containing the cDNA coding for the serine proteinase domain of NS3 under the control of the bacteriophage T7 gene 10 promoter. A solubilisation tag (ASKKKK) was inserted at the C terminus of the NS3 enzyme sequence (Steinkühler *et al.*, 1998). The ^{15}N, $^{15}N/^{13}C$ and $^{15}N/^{13}C/^{2}H$ uniformly labelled samples were obtained allowing the cells to grow in M9 minimum medium supplemented with 1 g/l [^{15}N]ammonium sulphate (Martek), 2.0 g/l [^{13}C]glucose (Martek). A further addition to the medium of 6.8 mg/l $ZnCl_2$ was necessary, since the protein is a zinc-binding proteinase. For the perdeuterated sample, growth was carried out in 99% $^{2}H_2O$ (Martek). The ^{15}N selectively labelled sample (Leu, Val and Ala) was obtained by incorporation of 0.33 g/l each of ^{15}N selectively labelled Leu, Val and Ala in M9 modified medium. The NMR samples were prepared by dialysis against a buffer containing 20 mM sodium phosphate, 4% deuterated glycerol (Isotec Inc.), 3 mM DTT, 1.5 mM Chaps (pH 6.3), 5 mM NaN_3. The protein concentration was in the range of 0.7-0.9 mM. The aggregation state of the samples was verified with dynamic light-scattering and sedimentation equilibrium studies. The solutions were monodispersed and the protein behaviour was compatible with a monomeric state in solution (S. Di Marco & M. Sollazzo, unpublished results).

Data collection and assignment

All the NMR experiments were acquired at 298 K on a Bruker AMX 500 MHz, Varian Unity Plus 600 MHz, Bruker DMX 600, and 800 MHz all equipped with z-shielded gradient triple resonance probes. Spectral assignments were obtained using the following 3D experiments: CT-HNCO, CT-HNCA, CT-HNCOCA, CBCACONH, HNCAHA, HN(COCA)HA, HACACO, (H)CCH-COSY, H(C)CH-COSY, H(C)CH-TOCSY, (H)CCH-TOCSY, C-CONH-TOCSY, ^{15}N edited ^{1}H TOCSY (Clore & Gronenborn, 1991a,b; Bax & Grzesiek, 1993). The following 3D experiments were acquired for the perdeuterated sample: CT-HNCA, CT-HNCOCA, HNCOCACB, HNCACB (Yamazaki *et al.*, 1994).

NOE-type 3D experiments were acquired (Clore & Gronenborn, 1991a): on the perdeuterated sample, ^{15}N edited NOESY (80 ms); on the double-labelled sample, two sets of ^{13}C edited NOESY (80, 100 and 150 ms), one

optimized for aliphatic and another set for aromatic residues; on the ^{15}N single-labelled sample, ^{15}N edited NOESY (80 and 150 ms) and ^{15}N edited ROESY (20 ms). On the ^{15}N selectively labelled Leu, Ala and Val sample, ^{15}N edited NOESY (100 ms) and TOCSY (24 ms) were also acquired.

Coupling constants and stereospecific assignments: on the ^{15}N sample, 3D experiment HNHA (Kuboniwa *et al.*, 1994); on the double labelled sample, 3D HAHB (Grzesiek *et al.*, 1994) experiment, and 2D CO and N decoupled (Hu *et al.*, 1997). On the 10% ^{13}C-labelled sample was acquired a CT-HSQC for the stereospecific assingment of methyl groups of Leu and Val (Neri *et al.*, 1989).

Spectra were processed using NMRPipe (Delaglio *et al.*, 1995) and analysed using NMRView (Johnson & Blevins, 1994) software packages.

Structure calculation

Approximate interproton distances were derived from the multidimensional NOE spectra (Clore & Gronenborn, 1991a). NOEs were grouped into three distance ranges 1.8-2.8 Å (1.8-3.0 Å for NOEs involving HN protons), 1.8-3.4 Å (1.8-3.6 Å for NOEs involving HN protons) and 1.8-5.0 Å (1.8-6.0 for NOEs involving methyl groups); 0.6 Å was added to the upper bounds of the strong and medium NOEs involving methyl groups. No constraint was included for the zinc-binding site during the early stages of the calculation. After verification that, in each structure, the ligands were always disposed approximately in a tetrahedral geometry, the zinc atom was subsequently incorporated into the calculations using six distance and one valence angle restraint involving the cysteine residues and the explicit zinc atom (Omichinski *et al.*, 1990). H149 was not constrained to bind the zinc atom, since our previous results suggested that the coordination could be mediated through a hydroxyl group (Urbani *et al.*, 1998). Protein backbone hydrogen-bonding restraints ($d_{NH\text{-}O} = 1.6\text{-}2.9$ Å, $d_{N\text{-}O} = 2.4\text{-}3.6$ Å) within areas of regular secondary structure were introduced during the final stages of refinement using standard NMR criteria based on backbone NOE $^3J_{HN\alpha H}$ coupling constants, supplemented with secondary ^{13}C shifts. The ϕ, ψ and $\chi 1$ torsion angle restraints were derived from homo- and heteronuclear three-bond coupling constant data, employing as minimum ranges ±25°, ±40° and ±30°, respectively. The structures were calculated by simulated annealing with the program X-Plor 3.851 (Brunger, 1993) on a SGI O2 R10000 platform, using a protocol described by Omichinski *et al.* (1997). Figures and statistical analysis were generated using the program InsightII (Molecular Simulations Inc.).

Brookhaven Protein DataBank

The coordinates of the final 20 simulated annealing structures have been deposited in the Brookhaven Protein DataBank, accession code 1bt7.

Acknowledgements

We thank Professor C. Griesinger, of the Frankfurt University Centre of Biomolecular NMR, (Frankfurt-am-Main) for access to the Large Scale Facility 800 and 600 MHz instruments, and Dr Marcus Maurer and Dr Teresa Carlomagno for the recording of the ^{13}C and ^{15}N NOESY at 800 MHz, and one of the ^{13}C edited H(C)CH TOCSY experiments on the 600 MHz. G.B. thanks all the people working at that facility for their kindness during his stay there. We thank Dr Bruce Johnson, (Merck Research Laboratories, New Jersey, USA) for his help in the collection of data on the perdeuterated sample. We are deeply indebted to Professor Stephan Grzesiek (Forschungszentrum, Jülich, Germany) for his help and recording of the experiments aimed at determining the dihedral angles.

References

Bachovchin, W. W. (1985). Confirmation of the assignment of the low field proton resonance of serine proteinases by using specifically nitrogen-15 labeled enzyme. *Proc. Natl Acad. Sci. USA*, 82, 7948-7951.

Bartenschlager, R. (1997). Molecular targets in inhibition of hepatitis C virus replication. *Ann. Chem. Chemother.* 8, 281-301.

Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Acct. Chem. Res.* 26, 131-138.

Bazan, J. F. & Fletterick, R. J. (1988). Viral cysteine proteases are homologuos to the trypsyn-like family of serine proteases: structural and functional implications. *Proc. Natl Acad. Sci. USA*, 85, 7872-7876.

Blevins, R. A. & Tulinsky, A. (1985). The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. *J. Biol. Chem.* 260, 4264-4275.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C. & Schmidt, G. (1983). Refined 2.0 Å X-ray structure of porcine pancreatic kallikrein A. *J. Mol. Biol.* 164, 237-282.

Bode, W., Walter, J., Huber, R., Wenzel, H. R. & Tschesche, H. (1984). The refined 2.2 Å (0.22 nm) X-ray crystal structure of the ternary complex formed by bovine trypsinogen, valine-valine and the Arg15 analogue of bovine pancreatic trypsin inhibitor. *Eur. J. Biochem.* 144, 185-190.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, R. S. & Hofsteenge, J. (1989). The refined 1.9 Å X-ray structure of human alpha-thrombin. *EMBO J.* 8, 3467-3475.

Brunger, A. T. (1993). *X-Plor Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Chang, J. Y. (1986). The structures and proteolytic specificities of autolysed human thrombin. *Biochem. J.* 240, 797-802.

Choi, H. K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. (1997). Structure of Semliki forest virus core protein. *Proteins: Struct. Funct. Genet.* 27, 345-359.

Choo, Q.-L., Kuo, G., Weiner, A. J., Bradley, L. R. D. W. & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood borne non-A non-B viral hepatitis genome. *Science*, 244, 359-362.

Cicero, D. O., Barbato, G., Koch, U., Ingallinella, P., Bianchi, E., Nardi, M. C., Steinkhüler, C., Cortese, R., Matassa, V., De Francesco, R., Pessi, A. & Bazzo, R. (1999). Structural characterization of the interactions of optimized product inhibitors with the protease domain of human hepatitis C virus by NMR and modelling studies. *J. Mol. Biol.*

Clore, G. M. & Gronenborn, A. M. (1991a). Structures of larger proteins in solution: three and four dimensional heteronuclear NMR spectroscopy. *Science*, 252, 1390-1399.

Clore, G. M. & Gronenborn, A. M. (1991b). Applications of three and four dimensional heteronuclear NMR

spectroscopy to protein structure determination. *Prog. NMR Spectrosc.* **23**, 43-92.

Coombs, G. S., Dang, A. T., Madison, E. L. & Corey, D. R. (1996). Distinct mechanisms contribute to stringent substrate specificity of tissue-type plasminogen activator. *J. Biol. Chem.* **271**, 4461-4467.

De Francesco, R., Urbani, A., Nardi, M. C., Tomei, L., Steinkühler, C. & Tramontano, A. (1996). A zinc site in viral serine proteinases. *Biochemistry*, **35**, 13282-13287.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

Failla, C., Tomei, L. & De Francesco, R. (1995). An amino-terminal domain of the hepatitis C virus NS3 proteinase is essential for interaction with NS4A. *J. Virol.* **69**, 1769-1777.

Fersht, A. (1984). *Enzyme Structure and Mechanism*, 2nd. edit., W. H. Freeman, San Francisco, CA.

Frey, P. A., Whitt, S. A. & Tobin, J. B. (1994). A low-barrier hydrogen bond in the catalytic triad of serine proteinases. *Science*, **82**, 1927-1930.

Fujinaga, M. & James, M. N. G. (1987). Rat submaxillary gland serine proteinase tonin structure resolution and refinement at 1.8 Å resolution. *J. Mol. Biol.* **195**, 373-396.

Fujinaga, M., Delbaere, T. J., Brayer, G. D. & James, M. N. G. (1985). Refined structure of α-lytic proteinase at 1.7 Å resolution. *J. Mol. Biol.* **183**, 479-502.

Greer, J. (1990). Comparative modeling methods: Application to the family of the mammalian serine proteinases. *Proteins: Struct. Funct. Genet.* **7**, 317-334.

Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. (1994). Multiple quantum line narrowing for measurements of Hα-Hβ *J* couplings in isotopically enriched proteins. *J. Am. Chem. Soc.* **117**, 5312-5315.

Houghton, M. (1996). Hepatatis C virus. In *Fields Virology* (Fields, B. N., Knipe, D. M. & Howley, P. M., eds), 3rd. edit., pp. 1035-1058, River Press, New York.

Hu, J. S., Grzesiek, S. & Bax, A. (1997). Two dimensional NMR methods for determining χ1 angles of aromatic residues in proteins from three bond $J_{C'C\gamma}$ and *J* $N_{C\gamma}$ couplings. *J. Am. Chem. Soc.* **119**, 1803-1804.

Ingallinella, P., Altamura, S., Bianchi, E., Taliani, M., Ingenito, R., Cortese, R., De Francesco, R., Steinkühler, C. & Pessi, A. (1998). Potent inhibitors of human hepatitis C virus NS3 proteinase are obtained by optimising the cleavage products. *Biochemistry*, **37**, 8906-8914.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. & Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 proteinase domain complexed with a synthetic NS4A cofactor peptide. *Cell*, **87**, 343-355.

Kuboniwa, H., Grzesiek, S., Delaglio, F. & Bax, A. (1994). Measurement of HN-H alpha *J* coupling in calcium free calmodulin using new 2D and 3D water flip back methods. *J. Biomol. NMR*, **4**, 871-878.

Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redecker, A. G., Purcell, R. H., Myamura, T., Dienstag, J. L., Alter, M. J., Syevens, C. E., Tagtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shister, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). An assay for circulating antibodies to a major etiologic virus human non-A non-B hepatitis. *Science*, **244**, 362-364.

Johnson, B. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603-614.

Landro, J. A., Raybuck, S. A., Luong, Y. P. C., O'Malley, E. T., Haberson, S. L., Morgenstern, K. A., Rao, G. & Livingston, D. J. (1997). Mechanistic role of an NS4A peptide cofactor with the truncated NS3 proteinase of hepatitis C virus: elucidation of the NS4A stimulatory effect via kinetic analysis and inhibitor mapping. *Biochemistry*, **36**, 9340-9348.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallog.* **26**, 283-291.

Lesk, A. & Fordham, W. D. (1996). Conservation and variability in the structures of serine proteinases of chymotrypsin family. *J. Mol. Biol.* **258**, 501-537.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T. & Hostomska, Z. (1996). The crystal strcture of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell*, **87**, 331-342.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., Margosiak, S., Dagostino, E. & Hostomska, Z. (1998). The conformation of hepatitis C virus NS3 proteinase with and without NS4A: a structural basis for the activation of the enzyme by its cofactor. *Clin. Diagnost. Virol.* **10**, 151-156.

Markley, J. L. (1978). Hydrogen bonds in serine proteinases and their complexes with protein proteinase inhibitors. *Biochemistry*, **17**, 4648-4656.

Meyer, E., Cole, G. & Radhakrishnan, R. (1988). Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Crystallog. sect. B*, **44**, 26-38.

Moult, J., Sussman, F. & James, M. N. G. (1985). Electron density calculation as an extension of protein structure refinement, *Streptomyces griseus* proteinase A at 1.5 Å resolution. *J. Mol. Biol.* **182**, 555-566.

Neddermann, P., Tomei, L., Steinkühler, C., Gallinari, P., Tramontano, A. & De Francesco, R. (1997). The non structural proteins of the hepatitis C virus: structure and functions. *Biol. Chem.* **378**, 469-476.

Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989). Specific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry*, **28**, 7510-7516.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three dimensional structures of proteins from interproton distance data by hybrid distance geometry simulated annealing calculations. *FEBS Letters*, **229**, 317-324.

Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990). High resolution three dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry*, **29**, 9324-9334.

Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M. & Clore, G. M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nature Struct. Biol.* **4**, 122-132.

Pizzi, E., Tramontano, A., Tomei, L., La Monica, N., Failla, C., Sardana, M., Wood, T. & De Francesco, R. (1994). Molecular model of the specificity pocket of the hepatitis C virus proteinase: implications for substrate recognition. *Proc. Natl Acad. Sci. USA*, **91**, 888-892.

Polgar, L. (1989). *Mechanisms of Proteinase Action*, CRC Press, Boca Raton, FA.

Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* proteinase B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry*, 22, 4420-4433.

Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. A. & Neurath, H. (1988). The structure of rat mast cell proteinase II at 1.9 Å resolution. *Biochemistry*, 27, 8097-8105.

Simmonds, P. (1994). Variability of hepatitis C virus. In *Current Studies in Hematology and Blood Transfusion* (Reesink, H. W., ed.), pp. 12-35, Kargel, Basel, Switzerland.

Steinkühler, C., Biasol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A. & De Francesco, R. (1998). Product inhibition of the hepatitis C virus NS3 proteinase. *Biochemistry*, 37, 8899-8905.

Tong, L., Wengler, G. & Rossmann, M. G. (1993). Refined structure of Sindbis virus core protein and comparison with other chymotripsin-like serine proteinase structures. *J. Mol. Biol.* 230, 228-247.

Urbani, A., Bazzo, R., Nardi, M. C., Cicero, D. O., De Francesco, R., Steinkühler, C. & Barbato, G. (1998). The metal binding site of the hepatitis C virus NS3 proteinase. *J. Biol. Chem.* 273, 18760-18769.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994). A suite of triple resonance experiments for the backbone assignments of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655-11666.

Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C. & Chen, Z. (1998). Complex of NS3 proteinase and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci.* 7, 837-847.

Yasutake, A. & Powers, J. C. (1981). Reactivity of human leukocyte elastase and porcine pancreatic elastase toward peptide 4-nitroanilides containing model desmosine residues. *Biochemistry*, 20, 3675-3679.

Edited by P. E. Wright

(Received 15 December 1998; received in revised form 19 March 1999; accepted 22 March 1999)

VIROLOGY 246, 104–112 (1998)
ARTICLE NO. VY989184

Critical Point Mutations for Hepatitis C Virus NS3 Proteinase

Kazunori Yamada,* Akiko Mori,*,[1] Makoto Seki,* Junko Kimura,* Satoshi Yuasa,*
Yoshiharu Matsuura,† and Tatsuo Miyamura†

*Molecular Medicine Laboratory, Yokohama Research Center, Mitsubishi Chemical Corp., 1000 Kamoshida-cho, Aoba-ku, Yokohama 227; and †Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162, Japan

Received January 15, 1998; returned to author for revision February 25, 1998; accepted April 13, 1998

The hepatitis C virus NS3 proteinase plays an essential role in processing of HCV nonstructural precursor polyprotein. To detect its processing activity, we developed a simple *trans*-cleavage assay. Two recombinant plasmids expressing the NS3 proteinase region and a chimeric substrate polyprotein containing the NS5A/5B cleavage site between maltose binding protein and protein A were co-introduced into *Escherichia coli* cells. The proteinase processed the substrate at the single site during their polyprotein expression. Deletion analysis indicated that the functionally minimal domain of the NS3 proteinase was composed of 146 amino acids, 1059 to 1204. We isolated several cDNA clones encoding the functional domain of the NS3 proteinase from the sera of patients chronically infected with HCV and determined their proteinase activity by this *trans*-cleavage assay. Both active and inactive clones existed in the same patients. Comparative sequence analyses of these clones suggested that certain point mutations seemed to be related to the loss of proteolytic activity. This was confirmed by back mutation experiments. Among the critical mutations, Pro-1168 to Thr and Arg-1135 to Gly were intriguing. These amino acids, which are situated near the oxyanion hole, seem to be essential for maintaining the conformation of the active center of the NS3 proteinase.

INTRODUCTION

Hepatitis C virus (HCV) is the major etiological agent of posttransfusion non-A, non-B hepatitis worldwide (Choo *et al.*, 1989; Kuo *et al.*, 1989). HCV infection results in mild and acute liver disease, but chronic infections are common and may eventually develop into cirrhosis or hepatocellular carcinoma (Saito *et al.*, 1990). Although interferons are currently used for the treatment of chronic hepatitis, their efficacy is limited to a small portion of patients owing to insufficient suppression of HCV replication. Therefore, another reliable anti-HCV agent is necessary to control HCV hepatitis.

HCV has positive-strand RNA approximately 9400 nucleotides long which encodes a single polyprotein of about 3010 amino acids (aa) (Choo *et al.*, 1989, 1991; Kato *et al.*, 1990; Takamizawa *et al.*, 1991). Since its genomic organization is similar to those of flaviviruses and pestiviruses, HCV is classified as a member of the family Flaviviridae (Miller and Purcell, 1990; Takeuchi *et al.*, 1990). Following the 5′-untranslated region, the viral structural proteins are located at the amino (N)-terminal region of the polyprotein in the order of core, E1 and E2. After being translated as a precursor polyprotein, the structural proteins are processed by a host cell signal peptidase(s), perhaps with no apparent involvement of virus-coded proteinase. The nonstructural (NS) proteins, which represent the essential machinery for viral replication, are located in the carboxyl (C)-terminal region in the order of NS2-NS3-NS4A-NS4B-NS5A-NS5B (Rice, 1996). In contrast to the cleavage of the structural proteins by the host enzyme, the NS proteins are processed by virus-coded proteinases. The HCV has two proteinases, NS2/3 proteinase and NS3 serine proteinase. The cleavage at NS2-NS3 is mediated by the former proteinase encoded within a region composed of the C-terminal portion of the NS2 gene and the N-terminal portion of the NS3 gene. Most likely this enzyme is a zinc-dependent metalloproteinase whose His-952 and Cys-993 were involved in catalysis (Grakoui *et al.*, 1993a; Hijikata *et al.*, 1993a). This putative metalloproteinase partially overlaps with the NS3 serine proteinase and presumably autocleaves the NS2/NS3 junction. The NS3 serine proteinase was shown to cleave the NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B junctions (Manabe *et al.*, 1994). The active site of this proteinase is composed of three highly conserved aa residues, His-1083, Asp-1107, and Ser-1165, which are well known as a catalytic triad of the serine proteinase family (Bazen and Fletterick, 1990; Bartenschlager *et al.*, 1993). It was confirmed that the Ser-1165 in the NS3 protein was essential for cleaving the downstream portion of the polyprotein by using *in vitro* transcription-translation systems and some mammalian cell culture systems (Chambers *et al.*, 1990; Gra-

Sequence data from this article have been deposited with GenBank under Accession Nos. AB013620–AB013651

[1] To whom reprint requests should be addressed. Fax: 81-45-963-3967. E-mail: fruits@atlas.rc.m-kagaku.co.jp

0042-6822/98 $25.00

koui *et al.*, 1993b,c; Hijikata *et al.*, 1993b; Tomei *et al.*, 1993). Comparison of the sequences around each cleavage site revealed the unique substrate specificity of the proteinase: Cys or Thr at the P1 position, Ser or Ala at the P1′ position, and Asp or Glu at the P6 position (Grakoui *et al.*, 1993b). Among other NS proteins, NS4A, a 54-residue amphipathic peptide, was shown to act as a cofactor for the NS3 proteinase interacting with the N-terminal portion of the enzyme (Bartenschlager *et al.*, 1995). Recently, the crystal structures of NS3 serine proteinase were reported from two groups. The tertiary structure of the proteinase was revealed to adopt a chymotrypsin-like folding, and its unique conformational aspects including a zinc-binding site and its complex formation with an NS4A peptide were elucidated (Kim *et al.*, 1996; Love *et al.*, 1996).

Since the viral serine proteinase is an attractive target for antiviral therapy, several assay systems of NS3 proteinase using *in vitro* transcription–translation systems and some mammalian cell culture systems have already been developed. Previously, we constructed an enzymatic assay system using one of our original HCV clones, D51, which displays functional activity of the purified enzyme *in vitro*, and studied the characteristics of the proteinase (Mori *et al.*, 1996). We also reported the importance of the N-terminal part of NS3 for its proteinase activity (Mori *et al.*, 1997). In parallel with these previous studies, we newly cloned HCV proteinase genes from HCV-positive sera of chronic hepatitis patients into a plasmid of *Escherichia coli* and examined their ability to cleave a coexpressed substrate fusion protein containing the NS5A/5B cleavage site between maltose binding protein (MBP) and protein A. It was found that the isolated NS3 proteinase clones had variety of activities and some of them were inactive. In this study, we identified critical point mutations at Pro-1168 and Arg-1135, which lead to loss of proteinase activity, and discuss how these amino acids contribute to the processing activity from a structural point of view.

RESULTS

Construction of *trans*-cleavage assay system in *E. coli*

To evaluate the activity of isolated HCV NS3 proteinase clones, we developed an assay system which detects the proteinase activity by cotransformation of *E. coli* with expression plasmids containing a cloned NS3 pro-

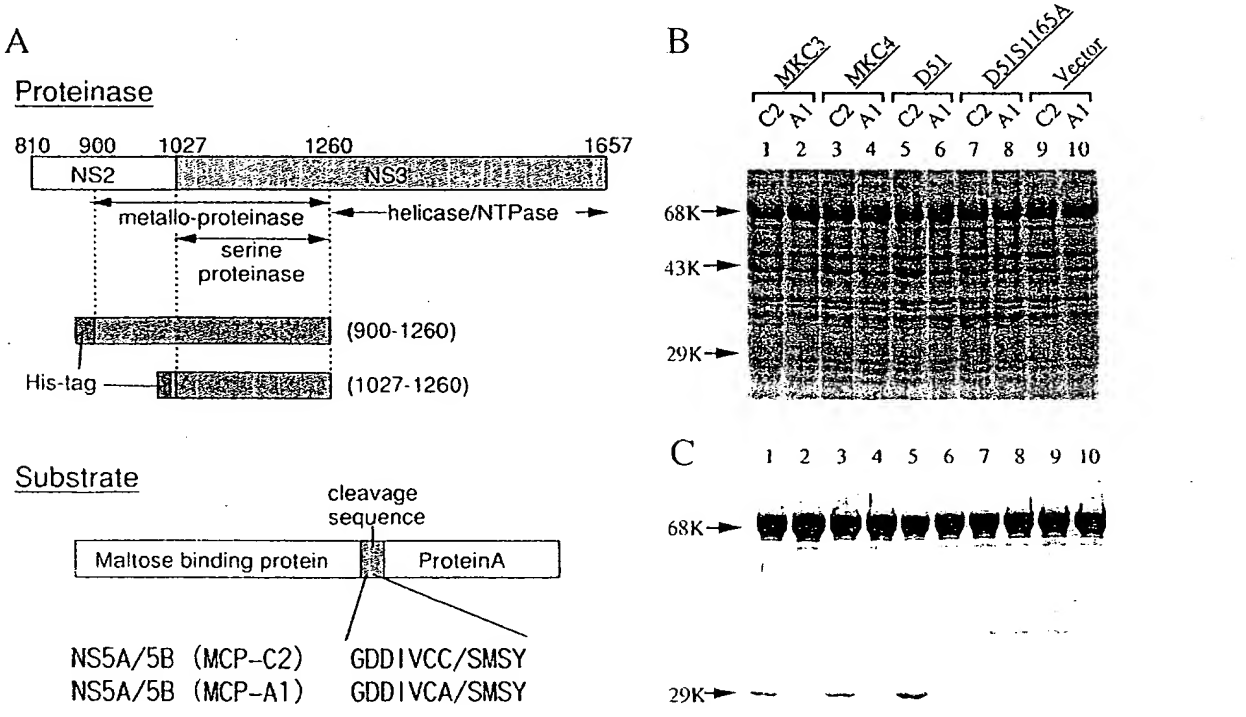


FIG. 1. Construction of *trans*-cleavage assay system in *E. coli*. (A) Schematic representation of the recombinant substrates and the HCV NS3 proteinases produced by the expression plasmids. (B) A substrate (MCP-C2, lanes 1, 3, 5, 7, and 9; MCP-A1, lanes 2, 4, 6, 8, and 10) and an enzyme clone (MKC3, lanes 1 and 2; MKC4, lanes 3 and 4; D51, lanes 5 and 6; D51S1165A, lanes 7 and 8; vector only, lanes 9 and 10) were coexpressed in double transformants of the substrate and the enzyme expression vectors by IPTG induction. The proteolytic activity was analyzed by SDS–PAGE followed by CBB staining. The substrate (68 kDa) and the products (43 and 29 kDa) are indicated by arrows. (C) Western blotting of the same sample as (B) using HRP-conjugated goat IgG anti-rabbit IgG. Only the substrate and the processed C-terminal product (68 and 29 kDa), which contained a protein A domain, were detected. The details are given under Materials and Methods.

teinase and a recombinant substrate. First, our proteinase-active original cDNA clones pMKC3, pMKC4, and pD51, which encode the residues 900–1260 of the viral polyprotein, were introduced into *E. coli* together with substrate expression plasmid pMCP-C2, which encodes a fusion protein containing the NS5A/5B cleavage sequence between MBP and protein A junctions (Fig. 1A). As shown in Fig. 1B, we found at the CBB staining level that the substrate (68 kDa) was cleaved into two polypeptides, MBP (43 kDa) and protein A (29 kDa). The processing activity was confirmed by Western blotting analysis using HRP-conjugated IgG (Fig. 1C). N-terminal sequence analysis of the protein A band produced by the proteinase revealed that the substrate was indeed processed at the NS5A/5B cleavage site (data not shown). When the Ser-1165 of the enzyme was replaced with Ala (Figs. 1B and 1C, lanes 7) or the P1 Cys residue of the recombinant substrate was replaced with Ala (Figs. 1B and 1C, lanes 2, 4, 6, 8, and 10), the cleavage was not observed. These results indicated that this assay is useful for screening for functional cDNA clones with proteinase activity.

Deletion analysis of the proteinase region

Using this *E. coli trans*-cleavage assay, we then determined the minimal region to maintain the proteinase activity. A series of N- and C-terminal deletion mutants from the D51 clone were constructed and subjected to the *trans*-cleavage assay (Figs. 2A–2C). Although the cleavages by the proteinase region (900–1260) and (1059–1214) were not efficient for detection by CBB staining (Fig. 2B), they were clearly detected by Western blotting (Fig. 2C). The N-terminal deletion experiments (lanes 2, 5, 7–14) indicated that the N-terminal border essential for *trans* cleavage of the substrate at the 5A/5B site was Val-1059. On the other hand, the C-terminal deletion experiments (lanes 2–6, 15) indicated that the functional C-terminal border was Thr-1204. From these results, the minimal NS3 proteinase region was narrowed down to the region between aa residues 1059 and 1204. To detect the proteinase activity sensitively, we adopted Western blotting analysis in the following experiments.

cDNA cloning of NS3 proteinase region from a patient and determination of its activity

Using RT–PCR methods, we isolated cDNA fragments of the HCV genome coding NS3 proteinase region 1027–1260 from two patients' sera (N and U). From each serum, we obtained several clones and examined them individually for their proteinase activity (Fig. 3). Several clones did not cleave the substrate (Fig. 3, clones MKC2, N-A1,

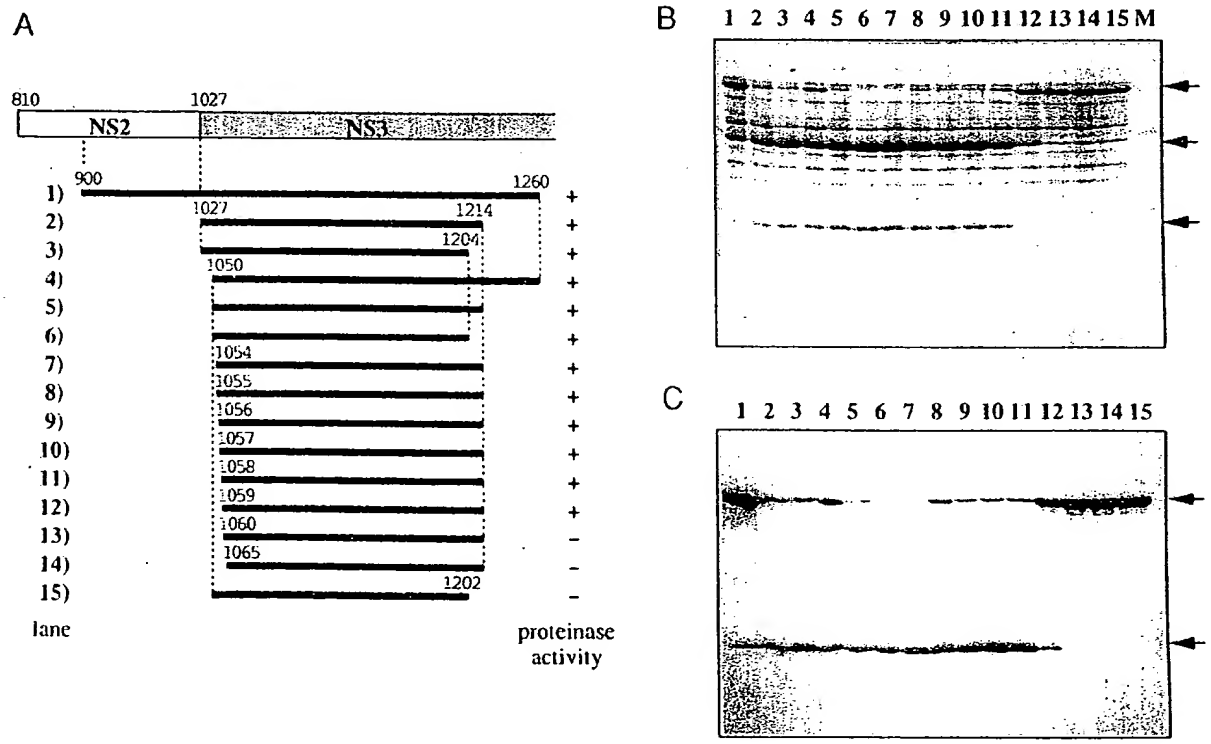


FIG. 2. Mapping of the minimal NS3 region required for cleavage at the NS5A/NS5B site. A series of N- and C-terminal deletion mutants from D51 proteinase clone were constructed and subjected to the *trans*-cleavage assay. (A) Schematic representation of the deletion mutants used in the assay. The numbers indicated on the left side correspond to the lane numbers in (B) and (C). (B) The processing activities of each deletion mutant were determined by SDS–PAGE followed by CBB staining. (C) The substrate and the processed C-terminal product, which contained a protein A domain, were detected by Western blotting using HRP-conjugated goat IgG anti-rabbit IgG.

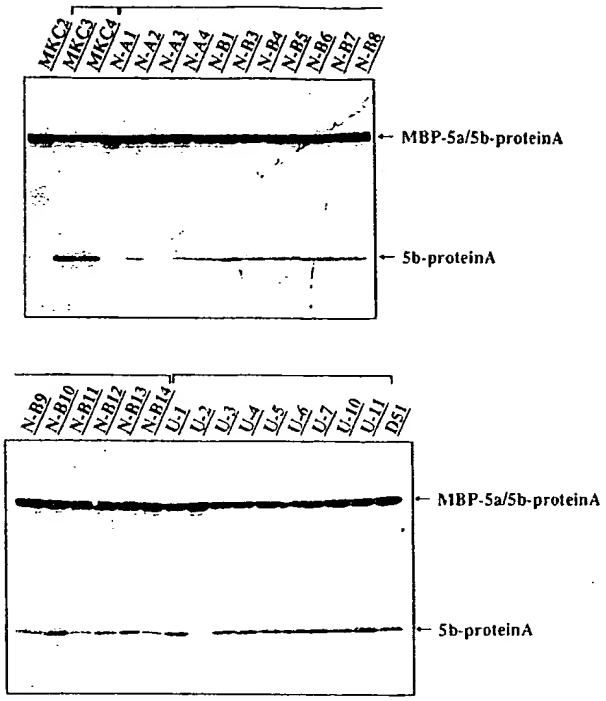


FIG. 3. Determination of processing activities of NS3 proteinase cDNA clones obtained from HCV-infected patients' sera. HCV NS3 proteinase cDNAs encoding the aa sequence from 1027 to 1260 of the HCV precursor polyprotein were cloned into the expression plasmid as described under Materials and Methods. Each clone was coexpressed with the substrate expression plasmid pMCP-C2 (Fig. 1A) in *E. coli* and its processing activity was determined. Both precursor and processed polyproteins containing a fused protein A domain were detected by Western blotting with HRP-conjugated IgG. Clones N-A1–A4 and N-B1–B14 were isolated from the serum of patient N. Also, clones U-1–11 were isolated from another sole serum source, patient U.

N-A3, and U-2). Since the expression of the NS3 enzyme was detected in each assay sample except for clone N-A1 (data not shown), the cDNA clones MKC2, N-A3, and U-2 were considered not functional.

To study the correlation between their aa sequences and their cleavage activities, we analyzed DNA sequences of each isolated clone. In the total nucleotide sequences, 1.18 and 0.27% of base substitutions from the consensus sequence were found in the clones derived from sera N and U, respectively (data not shown). These mutation rates account for 0.85% of the combined total sequences. However, 85% of the substitutions occurred at the third codon and aa changes were not frequent. Figure 4 shows aa sequences of each clone. The sequences of clone N-B5, N-B6, N-B9, N-B11, and N-B13 were the same as that of N-A2, and the sequence of N-B12 was the same as that of N-B3. Similarly, sequences of U-3, U-6, U-7, and U-11 were identical to that of the U-1 clone. From the comparative alignment, it was revealed that the sequences could be classified into two groups by their serum source, N or U at six positions (at 1115, Pro and Ser; at 1148, Asn and Thr; at 1151, Ser and Ala; at 1196, Val and Ile; at 1222, Ala and Thr; at 1239, Lys and Arg, respectively). In addition to such diversity, each clone had a few extra minor point mutations. However, the majority of such minor mutations did not affect proteinase activity (Fig. 3). Among four inactive clones, clone N-A1 had a single base pair deletion which made a stop codon at aa 1079 (Fig. 5A). The appearance of a stop codon was consistent with the finding that the enzyme expression was not detected in clone N-A1 (data not shown). The aa sequence of its back mutant N-A1N, instead of the sequence of N-A1, is shown in Fig. 4. The other inactive clones, MKC2, N-A3, and U-2, had a replacement of Arg-1135 by Gly, Pro-1168 by Thr, and His-1083 by Leu, respectively. These point mutations seemed to cause the loss of activity of the NS3 enzyme.

Analysis of the inactive proteinase clones

The comparative alignment revealed that the aa sequence of clone U-2 was identical to that of clone U-1 except for His-1083. Therefore, it was concluded that the loss of the proteinase activity was due to the point mutation at His-1083 to Leu. Since His-1083 is one of the catalytic triad residues (His-1083, Asp-1107, Ser-1165), it was clear that the point mutation at this site caused the loss of proteinase activity (Hijikata *et al.*, 1993a).

To examine whether the aa substitutions in other inactive clones caused the loss of activity, we next constructed a back mutant for each inactive clone. A corresponding nucleotide was inserted to correct the nonsense mutation in clone N-A1 and a single base pair substitution was introduced in MKC2 and N-A3, respectively (Fig. 5A). When the back mutant clones N-A1N, MKC2N, and N-A3N were introduced in *E. coli* together with the substrate expressing plasmid, cleavages of the recombinant substrate were observed in all cases (Fig. 5B). These results confirmed that each predicted aa change caused the loss of proteinase activity. The immunoblot analysis using anti-NS3 polyclonal antibody showed that each enzyme, except for the clone N-A1, normally expressed in *E. coli* (Fig. 5B, bottom), eliminating the possibility that the failure of detection of the proteinase activity was due to the enzyme instability or low-level expression. To further analyze the contribution of positive charge at the side chain of position 1135 to the processing activity, we prepared additional mutants from clone D51, in which Arg-1135 was substituted by Lys (D51R1135K) or Gln (D51R1135Q) as well as Gly (D51R1135G). Interestingly, it was found that Gln as well as Lys could substitute for Arg. However, when the residue was changed to Gly, the D51 clone lost its activity, like the MKC2 clone (Fig. 5B). This finding indicates that any factor other than a positive charge at position 1135 may contribute to the NS3 proteinase activity. Furthermore, mobility shift of the enzyme was observed in the mutants at position 1135, suggesting the importance of this aa position for enzyme conformation (Fig. 5B, bottom).

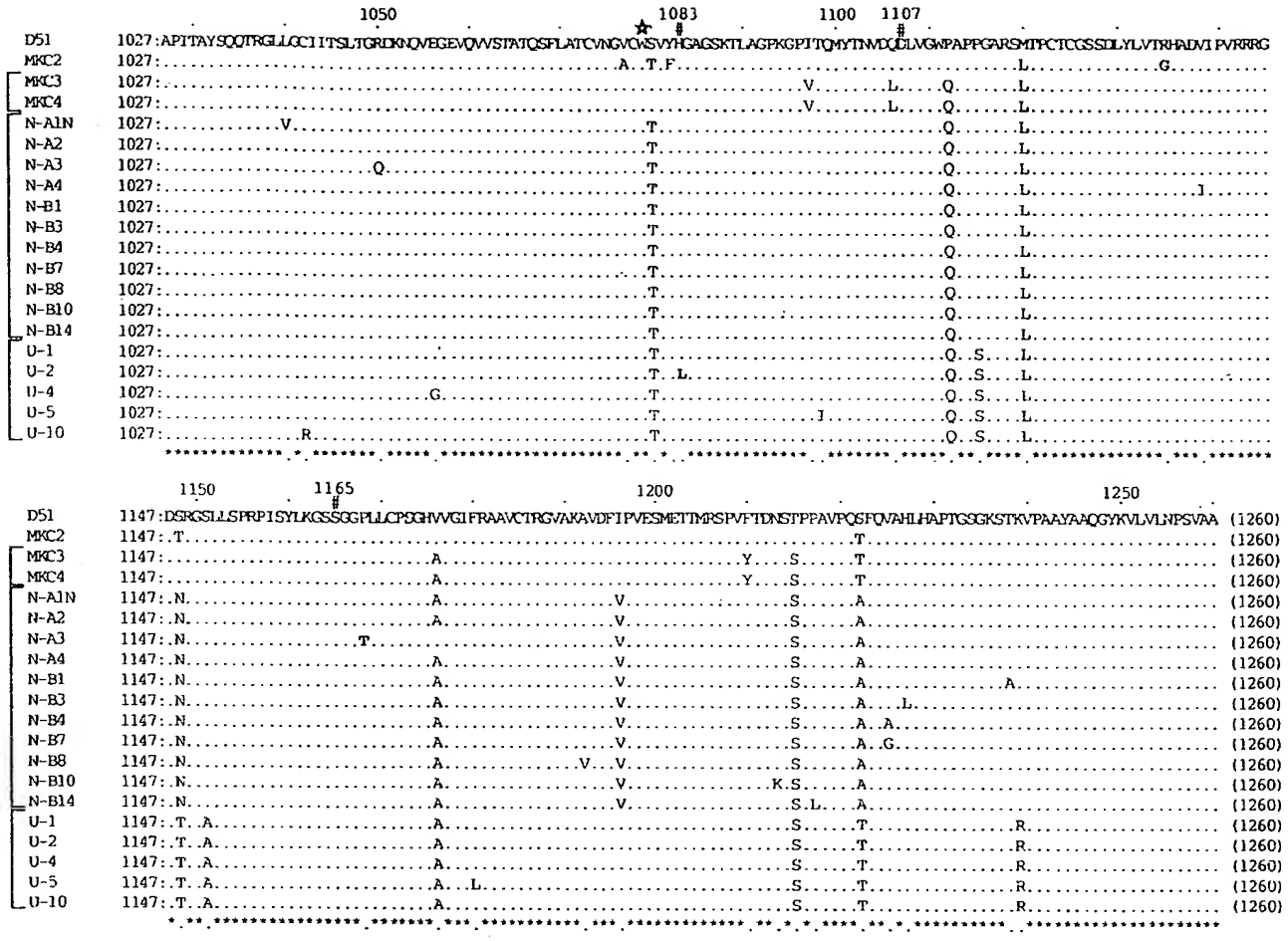


FIG. 4. Comparative alignment of the deduced amino acid sequences of the NS3 proteinase genes. MKC3, MKC4, and N-A1–A4 were initially obtained as 1196-bp DNA fragments encoding the aa sequence from 900 to 1260 of the HCV precursor polyprotein. The others were obtained as 716-bp DNA fragments encoding the aa sequence from 1027 to 1260 of the HCV precursor polyprotein. Although MKC3 and MKC4 are identical within the NS3 region, they are different within the NS2 region. The clones isolated from the same patient's serum are indicated by brackets. N-A1N is a proteinase-positive revertant of N-A1, which was mutated by a single base pair (G/C) insertion at the starred position (Fig. 5). Amino acid residues predicted to be null mutations are indicated in bold capitals. The residues predicted to form a catalytic triad (H1083, D1107, and S1165) are indicated by #.

DISCUSSION

We developed a *trans*-cleavage assay of HCV NS3 proteinase which detected its specific cleavage activity in *E. coli*. By using a coexpressed proteineous substrate, the cleavage activity was easily detected by SDS–PAGE and Western blotting. Using this *trans*-cleavage assay, we determined the minimal proteinase region in the N-terminal third of the NS3 gene by constructing a series of deletion mutants of the D51 clone. Previously, the proteinase domain was identified within region 1049–1215 (Tanji *et al.*, 1994) and we also used the region 1050–1214 as an active NS3 proteinase for its characterization (Mori *et al.*, 1996). In this study, the analyses of additional deletion mutants revealed that the minimal proteinase region was mapped between Val-1059 and Thr-1204 of HCV precursor protein, which was constructed with 146 aa residues (Fig. 2). The tertiary structure of the proteinase was revealed to adopt a chymotrypsin-like folding (Kim *et al.*, 1996; Love *et al.*, 1996). It was reported that positions of secondary structure elements were well matched to those of chymotrypsin, although some β strands did not superimpose with the equivalent strands in other chymotrypsin-like proteinases. Both the terminal residues of the minimal region, Val-1059 and Thr-1204, are located at the N-terminal end of β strand A1 and the end of C-terminal α helix, respectively (Kim *et al.*, 1996; Love *et al.*, 1996). NS3(1059–1204) covers the minimal region containing all secondary structures which are conserved among chymotrypsin-like proteinase family members. Our results suggest that this region forms a core domain essential for processing activity.

In the course of studying the activity of HCV proteinase cDNA clones obtained from HCV-infected patients, we found both functionally active and inactive clones of NS3 proteinase by the *trans*-cleavage assay (Fig. 3). It is interesting that some inactive clones (N-A1, N-A3 from serum N and U-2 from serum U) existed together with

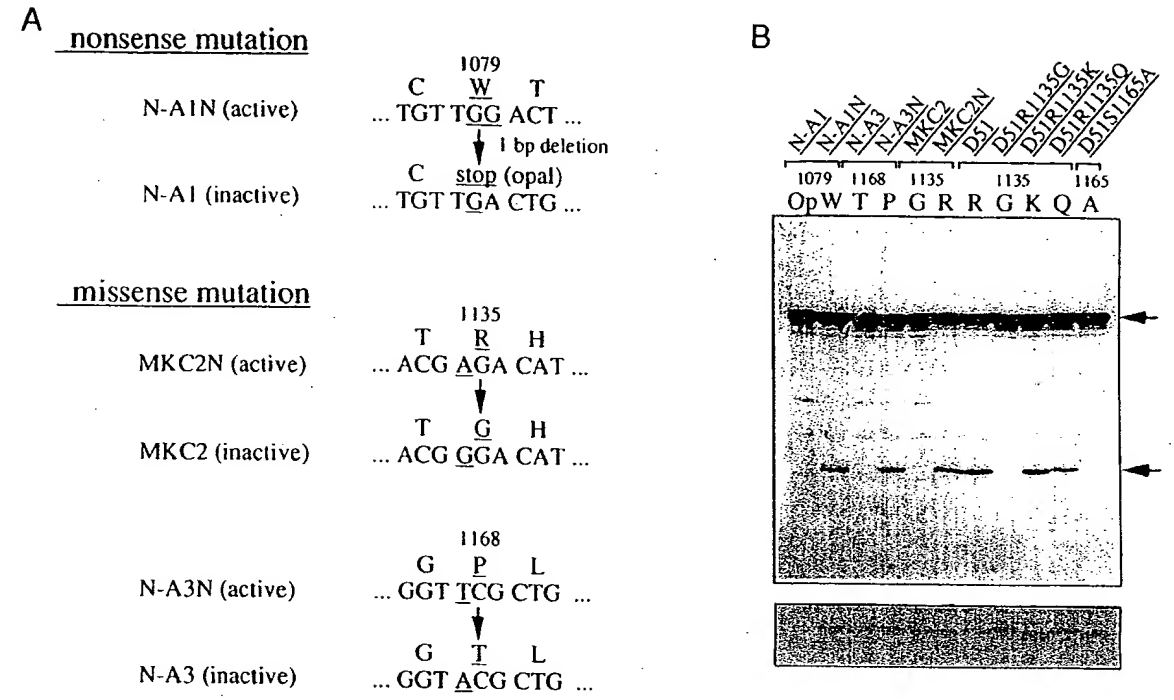


FIG. 5. Mutation analysis of proteinase-defective clones. (A) Predicted mutation sites of proteinase-defective clones. Underlines indicate replaced aa residues and nucleotides. (B) To confirm the predicted null mutation sites shown in (A), back mutation experiments were performed. The Opal mutation (indicated as Op in the figure) in N-A1 was repaired by single G/C base pair insertion. The aa sequence of the resultant clone designated as N-A1N is shown in Fig. 4. The Thr-1168 in N-A3 and the Gly-1135 in MKC2 were replaced by Pro (N-A3N) and Arg (MKC2N), respectively. Furthermore, the Arg-1135 in D51 (an active clone) was replaced by Gly (D51R1135G), Lys (D51R1135K), or Gln (D51R1135Q) residues. D51S1165A, in which the Ser-1165 was replaced by Ala residue, was used as inactive control. The proteinase activities of these clones were determined by *trans*-cleavage assay in *E. coli*. The bottom part of this figure shows the expression of NS3 proteinase analyzed by Western blotting using rabbit polyclonal anti-NS3 antibody.

active clones in the same patient's serum. The sequencing analysis of the cloned NS3 genes revealed that the base substitutions amounted to 0.85% in the total nucleotide sequences (data not shown). Among the substitutions, 85% of the cases occurred at the third codon, which did not change the aa sequences. Since the analysis of the sequence of HCV genome might inevitably involve the possibility of nucleotide change due to PCR error, we are not able to entirely exclude such possibility. However, these biased nucleotide mutations may reflect the characteristics of an error-prone RNA-dependent RNA polymerase and the lack of an associated repair mechanism in the viral replication, which may cause the quasispecies of the HCV genome within infected individuals (Bukh *et al.*, 1995). Among such quasispecies, some mutations may happen to change an aa residue(s) critical for proteinase activity. In this study, we identified four cases of such critical mutations in sera of patients chronically infected with HCV.

Among the four inactive clones, N-A1 was found to have a nonsense mutation at position 1079 (Fig. 5A). Since the translation of the HCV polyprotein may be terminated at the stop codon, N-A1 is considered to be a clone defective in HCV replication. In another inactive clone, U-2, a point mutation occurred at a histidine residue of the catalytic triad (His-1083) and directly made the proteinase inactive. In the third case of N-A3, mutation at Pro-1168 was revealed to induce inactivation of the enzyme. The proline residue follows the "oxyanion-stabilizing loop" (Leu-1161 to Ser-1165), which constitutes an active center. It is possible that this proline residue is structurally critical and its substitution for threonine distorted the conformation of the main chain within the loop and caused loss of the enzyme activity. In the fourth inactive clone, MKC2, Arg-1135 was replaced by glycine, which was located on a loop between β strands A2 and B2 in the C-terminal domain and was positioned just behind the oxyanion hole of the active center (Kim *et al.*, 1996; Love *et al.*, 1996). The side chain at position 1135 may contribute to the processing activity by certain interactions with amino acids constituting the "oxyanion-stabilizing loop." Since a glutamine residue could substitute for Arg-1135 as shown in Fig. 5B, a positive charge at this position may not be very important for the interactions. The volume of the side chain might be critical in interactions such as hydrogen bonds, and the smallest side chain of glycine might lead to loss of such interaction causing the enzyme to be inactive in the clone MKC2.

The *trans*-cleavage assay described here could be used for analyses to measure the activity of NS3 proteinase clones and to identify quasispecies of the proteinase

sequences. The system would be also useful for revealing the amino acids critical in inactive HCV clones. Accumulation of such information should be helpful for understanding the relationship between the structure and the activity of the enzyme.

MATERIALS AND METHODS

Construction of substrate expression vectors

The expression plasmid pMCP1, which encodes *E. coli* MBP and *Staphylococcus aureus* protein A in tandem, is a derivative of pMAL-c2 (New England BioLabs, Inc., MA) and pRIT2 (Pharmacia Biotech, Inc., Uppsala, Sweden). Plasmid pMCP1 also contains multiple cloning sites (*Hin*dIII to *Xba*I site of pUC19) between the MBP gene and the protein A gene. Substrate expression plasmids, pMCP-C2 and pMCP-A1, which encode fusion proteins containing appropriate aa sequences around the NS3 proteinase cleavage site between the MBP and protein A junctions (Fig. 1A), were constructed by inserting phosphorylated linkers into the *Hin*dIII–*Xba*I site of pMCP1. The synthetic oligonucleotide pairs used in these constructions are as follows: for plasmid pMCP-C2, 5′-AGCTTGGCGACGACATCGTCTGCTGCTCAATGTCCTACT-3′ and 5′-CTAGAGTAGGACATTGAGCAGCAGACGATGTCGTCGCCA-3′; for plasmid pMCP-A1,5′-AGCTTGGCGACGACATCGTCTGCGCCTCAATGTCCTACT-3′ and 5′-CTAGAGTAGGACATTGAGGCGCAGACGATGTCGTCGCCA-3′.

Cloning of NS3 proteinase genes

The expression plasmids pWNH71 and pWB298 each contain a 13 N-terminal amino acid leader sequence derived from a nitrile hydratase gene, which are driven under the control of the *trc* promoter. This leader sequence possesses a consecutive stretch of histidine (His) residues which allows the fusion protein to be purified with ease in a single step by metal chelating affinity chromatography.

HCV RNA molecules were purified from 50 μl of an HCV patient's serum (a gift from Dr. N. Hayashi) by using RNAzol B (Tel-Test, Inc., Texas) and subjected to cDNA synthesis, followed by the reverse transcription–polymerase chain reaction (RT–PCR). To amplify the NS3 region, the following oligonucleotides were used as nested PCR primers: forward primers, YH351 (5′-TTAAGCTTCTCGGTCCGCTCATGGTRCTCCARGC-3′) and YH357 (5′-TTAAGCTTGCGCCYATCACGGCCTAYTCCC-3′); backward primers, YH354 (5′-TAAGATCTARGCRGCAACRGAIGGRAGGACG-3′) and KY390 (5′-TAGGATCCTARGCRGCAACRGAIGGRAGGACG-3′) (Underlines denote the restriction sites used for subcloning, and nucleotides in bold type indicate the stop codon.) To amplify the cDNA fragment encoding aa residues 900 to 1260 in the HCV polyprotein, a primer set of YH351 and YH354 was used. To amplify the cDNA fragment encoding aa residues 1027 to 1260 in the HCV polyprotein, a primer set of YH357 and KY390 was used.

The *Hin*dIII–*Bgl*II fragment and the *Hin*dIII–*Bam*HI fragment of the PCR products were ligated into the *Hin*dIII–*Bgl*II site of pWNH71 and the *Hin*dIII–*Bam*HI site of pWB298, respectively. To screen enzymatically active NS3 proteinase clones, as described later, the ligated plasmid DNA and the substrate expression plasmid pMCP-C2 were co-introduced into *E. coli* DH5 cells. Since the substrate and the enzyme expression vectors have different replication origins and different antibiotic-resistant markers, both plasmids can be maintained under selective pressure. Ampicillin and kanamycin double-resistant colonies were selected as the transformants harboring the two plasmids, and the plasmid DNAs were subjected to restriction endonuclease analysis.

Detection of NS3 proteinase activity in *E. coli* expression system

Recombinant *E. coli* strains harboring the two plasmids were precultured overnight in LB broth (Difco) containing ampicillin (50 μg/liter) and kanamycin (30 μg/liter) at 30°C with shaking. The cultured cells were inoculated at 1/50 dilution into the same broth and incubated further at 37°C. When the OD_{650} reached 0.8, isopropyl-b-D-thiogalactopiranoside (IPTG) was added at the final concentration of 0.5 mM, and the culture was further incubated for 4 h. After the cultured cells were harvested and suspended in 0.85% NaCl, a portion of the suspension was mixed with Laemmli's sample buffer and analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE).

SDS–PAGE was carried out in a 14% acrylamide slab gel and the gel was stained with Coomassie brilliant blue (CBB). In addition, Western blot analysis was carried out after SDS–PAGE. The processed products from the C-terminal region of the substrate protein, containing protein A-fused polypeptide, was detected with horseradish peroxidase (HRP)-conjugated goat IgG anti-rabbit IgG. To confirm the expression of the NS3 protein in *E. coli* cells, rabbit anti-NS3 IgG prepared from a rabbit immunized with purified recombinant protein (aa 900 to 1260; expressed in *E. coli*) was used as the first antibody, and HRP-conjugated goat IgG anti-rabbit IgG as the second.

Mutagenesis

Site-specific mutagenesis was accomplished by using a PCR-based method (Ho *et al.*, 1989). All constructs were confirmed by sequencing analysis. The PCR primers used were as follows: Vector primers; forward primer, 5′-TGTGAGCGGATAACAATTTCGGATC-3′, and backward primer, 5′-TCGGCCGCCCGACTATCACCGCCC-3′; mutagenesis primers; for MKC2 (G1135→R): 5′-CTTTACCTG-

GTCACGAGACATGCTGATGTC-3′ and 5′-GACATCAGCATGTCTCGTGACCAGGTAAAG-3′; for N-A1 (amber → W1079; 1 G/C bp insertion): 5′-AACGGCGTGTGTTGGACTGTCTACCATGG-3′ and 5′-ACCATGGTAGACAGTCCAACACACGCCGTT-3′; for N-A3 (T1168 → P): 5′-GCTCTTCGGGTGGTCCGCTGCTTTGCC-3′ and 5′-GGCAAAGCAGCGGACCACCCGAAGAGC-3′; for D51 (R1135 → G): 5′-CTTTACTTGGTCACGGGACATGCTGATGTC-3′ and 5′-GACATCAGCATGTCCCGTGACCAAGTAAAG-3′; for D51 (R1135 → K): 5′-CTTTACTTGGTCACGAAACATGCTGATGTC-3′ and 5′-GACATCAGCATGTTTCGTGACCAAGTAAAG-3′; for D51 (R1135 → Q): 5′-CTTTACCTGGTCACGCAACATGCTGATGTC-3′ and 5′-GACATCAGCATGTTGCGTGACCAGGTAAAG-3′; for D51 (S1165 → A): 5′-TGAAGGGTTCCTGCGGTGGTCC-3′ and 5′-GGACCACCGCAGGAACCCTTCA-3′, forward and backward primers, respectively.

Construction of deletion mutants

All deletion mutants were constructed by the PCR. The *Hin*dIII–*Bgl*II fragment and the *Hin*dIII–*Bam*HI fragment of PCR products were ligated into the *Hin*dIII–*Bgl*II site of pWNH71. All constructs were confirmed by sequencing analysis.

ACKNOWLEDGMENTS

We thank Dr. N. Hayashi (Osaka University) for the HCV-positive sera. We also thank Dr. Y. Honda (IRLI, Keniya), Dr. T. Matsuzaki, and Dr. H. Kubodera (Mitsubishi Chemical Corp.) for useful discussions. This work was supported in part by grants from the Japanese Human Science Foundation.

REFERENCES

Bazan, J. F., and Fletterick, R. J. (1990). Structural and catalytic models of trypsin-like viral proteases. *Semin. Virol.* 1, 311–322.

Bartenschlager, R., Ahlborn-Lakke, L., Mous, J., and Jacobsen, H. (1993). Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J. Virol.* 67, 3835–3844.

Bartenschlager, R., Lohmann, V., Wilkinson, T., and Koch, J. O. (1995). Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *J. Virol.* 69, 7519–7528.

Bukh, J., Miller, R. H., and Purcell, R. H. (1995). Genetic heterogeneity of hepatitis C virus: Quasispecies and genotypes. *Semin. Liver Dis.* 15, 41–63.

Chambers, T. J., Hahn, C., Galler, R., and Rice, C. M. (1990). Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* 44, 649–688.

Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., and Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244, 359–362.

Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G., and Houghton, M. (1991). Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* 88, 2451–2455.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M., and Rice, C. M. (1993a). A second hepatitis C virus-encoded proteinase. *Proc. Natl. Acad. Sci. USA* 90, 10583–10587

Grakoui, A., Wychowski, C., Lin, C., Feinstone, S. M., and Rice, C. M. (1993b). Expression and identification of hepatitis C virus polyprotein cleavage products. *J. Virol.* 67, 1385–1395.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M., and Rice, C. M. (1993c). Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J. Virol.* 67, 2832–2843.

Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K., and Shimotohno, K. (1993a). Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J. Virol.* 67, 4665–4675.

Hijikata, M., Mizushima, H., Tanji, Y., Komoda, Y., Hirowatari, Y., Akagi, T., Kato, N., Kimura, K., and Shimotohno, K. (1993b). Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* 90, 10773–10777.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77, 51–59.

Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K. (1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc. Natl. Acad. Sci. USA* 87, 9524–9528.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87, 343–355.

Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redecker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W., and Houghton, M. (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* 244, 362–364.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., and Hostomska, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* 87, 331–342.

Manabe, S., Fuke, I., Tanishita, O., Kaji, C., Gomi, Y., Yoshida, S., Mori, C., Takamizawa, A., Yoshida, I., and Okayama, H. (1994). Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. *Virology* 198, 636–644.

Matsuura, Y., and Miyamura, T. (1993). The molecular biology of hepatitis C virus. *Semin. Virol.* 4, 297–304.

Miller, R. H., and Purcell, R. H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc. Natl. Acad. Sci. USA* 87, 2057–2061.

Mori, A., Yamada, K., Kimura, J., Koide, T., Yuasa, S., Yamada, E., and Miyamura, T. (1996). Enzymatic characterization of purified NS3 serine proteinase of hepatitis C virus expressed in *Escherichia coli*. *FEBS Lett.* 378, 37–42.

Mori, A., Yuasa, S., Yamada, K., Nagami, Y., and Miyamura, T. (1997). The N-terminal region of NS3 serine proteinase of hepatitis C virus is important to maintain its enzymatic integrity. *Biochem. Biophys. Res. Commun.* 231, 738–742.

Rice, C. M. (1996). Flaviviridae: The viruses and their replication. *In* "Virology," pp. 931–960. *In* (B. N. Fields, D. M. Knipe, and P. M. Howley, (Eds.). Raven Press, New York.

Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, T. Y., Koi, S., Onji, M., Ohta, Y. *et al.* (1990). Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA* 87, 6547–6549.

Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I., and Okayama, H. (1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol.* 65, 1105–1113.

Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I., and Miyamura, T. (1990). The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequences of two isolates derived from an experimentally infected chimpanzee and healthy human carriers. *J. Gen. Virol.* 71, 3027–3033.

Tanji, Y., Hijikata, M., Hirowatari, Y., and Shimotohno, K. (1994). Identification of the domain required for *trans*-cleavage activity of hepatitis C virus serine proteinase. *Gene* 145, 215–219.

Tomei, L., Failla, C., Santolini, E., De Francesco, R., and La Monica, N. (1993). NS3 is a serine proteinase required for processing of hepatitis C virus polyprotein. *J. Virol.* 67, 4017–4026.

Cell, Vol. 87, 331–342, October 18, 1996, Copyright ©1996 by Cell Press

The Crystal Structure of Hepatitis C Virus NS3 Proteinase Reveals a Trypsin-like Fold and a Structural Zinc Binding Site

Robert A. Love,* Hans E. Parge,*
John A. Wickersham,* Zdenek Hostomsky,*
Noriyuki Habuka,†‡ Ellen W. Moomaw,*
Tsuyoshi Adachi,†‡ and Zuzana Hostomska*
*Agouron Pharmaceuticals, Inc.
3565 General Atomics Court
San Diego, California 92121
†Japan Tobacco Inc.
Central Pharmaceutical Research Institute
1-1, Murasaki-cho, Takatsuki
Osaka 569
Japan
‡Center for Tsukuba Advanced Research Alliance
University of Tsukuba
Tsukuba 305
Japan

Summary

During replication of hepatitis C virus (HCV), the final steps of polyprotein processing are performed by a viral proteinase located in the N-terminal one-third of nonstructural protein 3. The structure of NS3 proteinase from HCV BK strain was determined by X-ray crystallography at 2.4 Å resolution. NS3P folds as a trypsin-like proteinase with two β barrels and a catalytic triad of His-57, Asp-81, Ser-139. The structure has a substrate-binding site consistent with the cleavage specificity of the enzyme. Novel features include a structural zinc-binding site and a long N-terminus that interacts with neighboring molecules by binding to a hydrophobic surface patch.

Introduction

Hepatitis C virus (HCV) is the major etiologic agent of human parenterally and community-acquired non-A, non-B hepatitis (Choo et al, 1989). Chronic HCV infection is a global disease, and the number of carriers is estimated to be about 300 million. Chronic infection may lead to the development of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (reviewed by Houghton, 1996). In Europe and Japan, the disease is more prevalent than either hepatitis B virus or human immunodeficiency virus infections. Protective vaccination is not available for HCV, and current treatments with interferon are successful only in a limited number of patients. Therefore, considerable attention has been focused in recent years on understanding HCV replication and obtaining structural information about essential HCV proteins.

The HCV virion has a positive-strand RNA genome that was cloned in 1989 (Choo et al., 1989). It is composed of about 9,400 nucleotides and contains a single large open reading frame encoding a polyprotein of 3010–3033 amino acid residues (Kato et al., 1990; Choo et al., 1991; Takamizawa et al., 1991). The genetic organization of HCV (reviewed by van Doorn, 1994) is similar to that of flavi- and pestiviruses, and it was classified as a separate genus of the family Flaviviridae. Sequence analysis of HCV isolates reveals that HCV exists in many distinct variants. A total of six major genotypes and at least 11 subtypes have been recognized (Simmonds et al., 1994).

The nonstructural (NS) proteins involved in replication of the HCV genome are released by the action of two proteinases: NS2-3 and NS3. NS2-3 proteinase is a zinc-dependent enzymatic activity that performs a single proteolytic cut to release the N-terminus of NS3 (Grakoui et al., 1993, Hijikata et al., 1993). The action of NS3 proteinase (NS3P), which resides in the N-terminal one-third of the NS3 protein, then yields all remaining nonstructural proteins: NS4A, NS4B, NS5A, and NS5B. The C-terminal two-thirds of the NS3 protein contain a helicase. While the functional relationship of these two domains is unknown, the separately expressed proteinase and helicase domains of NS3 exhibit their respective activities in vitro (Suzich et al., 1993; D'Souza et al., 1995; Steinkühler et al. 1996).

The N-terminal domain of NS3 has been found to contain the catalytic motif of a trypsin-like serine proteinase (Miller and Purcell, 1990). The positions of amino acids His-57, Asp-81, and Ser-139 (numbering from the start of NS3) are strictly conserved among all HCV-derived sequences; their relative order and spacing in the sequence correspond to the catalytic triad of the trypsin family. However, the NS3P exhibits several other features that are highly unusual for a trypsin-like proteinase: it is covalently attached to a helicase possessing NTPase activity, it requires a protein cofactor (NS4A), and displays sensitivity to divalent metal ions. Using in vitro transcription/translation systems, the NS4A protein (located immediately downstream of NS3 on the polyprotein) was shown to be required for cleavage at the NS3/NS4A, NS4A/NS4B, and NS4B/NS5A sites. The NS3/NS4A cleavage occurs rapidly and in *cis* (intramolecular event), while the others occur in trans (intermolecular events). NS4A also accelerates the rate of cleavage at the NS5A–NS5B junction (Bartenschlager et al., 1994; Failla et al., 1994; Lin and Rice, 1995; Koch et al., 1996). While NS4A may have several functions, it has been suggested that NS3 and NS4A proteins form a complex that is important for modulation of proteolytic activity (Hijikata et al., 1993; Lin et al., 1994).

The proper understanding of these unique features of the NS3P as a molecular target can be achieved only in the context of a high resolution atomic structure, which in turn would also accelerate design and development of new drugs to treat HCV infection. However, crystallization of NS3P has proven difficult owing to its poor solubility and tendency to aggregate. To circumvent this problem, we examined a series of NS3P domains of variable sizes, derived from different HCV strains. Here, we report the three-dimensional structure of the NS3P domain from the strain BK, which belongs to the HCV genotype 2b. HCV BK NS3P shows significant amino acid sequence identity with representative HCV genotypes: 89% with HCV H (genotype 1a), 96% with

HCV J (1b), 71% with HCV J6 (2a), 71% with HCV J8 (2b), and 78% with HCV 3A (3a). It shares 33% sequence identity with the NS3 proteinase from the recently discovered hepatitis G virus (Linnen et al., 1996). Crystals of HCV BK NS3P were obtained using a 189 amino acid recombinant fragment purified from Escherichia coli. This structure represents the first view of a processing enzyme from the flavivirus family and opens new possibilities for design of drugs targeting HCV replication.

Results and Discussion

Overall Structural Features

HCV NS3 proteinase (NS3P) is folded into two six-stranded β barrels, similar to those of trypsin-like serine proteinases (Figures 1A and 2A–2B). However, apart from the three-catalytic triad residues (His-57, Asp-81, Ser-139) of NS3P and the sequence of GXSGG at Ser-139, it shares virtually no sequence similarity with proteinases possessing a trypsin-like fold (Figure 1B). In our discussions, features of NS3P are described while it is compared with other classes of trypsin-like proteinases. These classes include cellular enzymes such as pancreatic proteinases (e.g., elastase; Meyer et al., 1988) and bacterial proteinases (e.g., α-lytic proteinase; Fujinaga et al., 1985), all having serine as the active-site nucleophile. Also included are viral proteinases with a nucleophile of either serine (e.g., Sindbis virus core protein; Choi et al., 1991) or cysteine (e.g., rhinoviral 3C proteinase; Matthews et al., 1994). These examples were chosen because they contain an uncharged S1 specificity pocket, which for elastase and α-lytic enzymes is also relatively small. NS3P is expected to have a small uncharged specificity site because the consensus amino acid at substrate P1 is either cysteine or threonine (Grakoui et al., 1993).

The number of amino acids at the NS3P N-terminus, before the first β barrel, is significantly larger than in most proteinases in their active states. These initial 30 residues extend away from the protein and form several β strands that interact with neighboring molecules (Figure 2). The possible significance of this interaction is discussed later. The existence of secondary structure in the N-terminus is reminiscent of the short β strand found in many cellular proteinases (e.g., residues 19–22 of elastase) but differs from the case of the α helix found in the picornaviral 3C proteinases.

NS3P displays a spatial arrangement of strands within its β barrels similar to other trypsin-like proteinases; however, the loops connecting these strands are relatively short (Figures 2 and 3). In this regard, NS3P parallels the economical use of amino acids found in Sindbis core protein (Figure 3B), or possibly even a viral 2A proteinase (the presumed trypsin-like cysteine proteinase upstream of 3C in picornaviruses and enteroviruses, but with unknown structure). The total number of residues comprising the two β barrel motifs (from the first β barrel strand to the end of the last β barrel strand) is about 140 for NS3P, only slightly larger than 133 in Sindbis, and similar to the predicted number of about 140 for 2A proteinases (Bazan and Fletterick, 1988); this number exceeds 170 residues in the cellular proteinases. As a consequence of this economy, several loops common in the cellular proteinases are absent in NS3P, such as the calcium-binding loop (βD1 to βE1), the autolysis loop (βA2 to βB2), and the so-called methionine loop (βB2 to βC2). In the bacterial and picornaviral proteinases, this latter loop (but in pancreatic proteinases, the βE1 to βF1 loop) is positioned such that it can interact with residues on the P side of the substrate (Figure 3B); the absence of a corresponding loop in NS3P is consistent with its apparent lack of substrate recognition over P2 to P5 (Bartenschlager et al., 1995; Grakoui et al., 1993).

After the final strand βF2, there is one turn of α helix (171–174) that closely matches the first turn of the C-terminal helix found in cellular proteinases. Following this turn, the polypeptide is disordered in two of three monomers in our asymmetric unit. In the monomer discussed here, the chain turns back toward βE1 to form a short antiparallel β-interaction between βE0 (181–182) and βE1b (75–76; see Figure 2); this interaction may be the result of crystal packing only. Questions of flexibility and conformation at the C-terminus of NS3P are relevant because this crystal structure represents an enzyme which in vivo is permanently attached (via its C-terminus) to a helicase, the two entities together defining the NS3 protein. The exact range of residues within NS3 that corresponds to the folded helicase is unknown, as is the length of the polypeptide linker between NS3P and the helicase.

Zinc-Binding Site

In cell-free transcription/translation experiments, processing of the HCV NS polyprotein by NS3P is stimulated by the addition of Zn^{2+}. Inductive coupled-plasma mass spectroscopy of purified recombinant NS3P reveals an equimolar ratio of NS3P and zinc (Z. H., unpublished data). Thus, the existence of a functionally relevant zinc in HCV NS3P was expected. From our NS3P homology models, the close proximity of conserved cysteines 97, 99, 145, and histidine 149 suggested a zinc-chelation site, because these residues are frequently members of structural zinc sites (Vallee and Auld, 1990; Schwabe and Klug, 1994), and because disulfide linkages are unlikely in an intracellular proteinase. The crystal structure confirms this idea; these three cysteines together provide a partial tetrahedral geometry around the zinc ion (see Figure 2B), with cysteine sulfur to zinc distances of 2.0–2.5 Å. In two of three monomers in the asymmetric unit of the crystal, the fourth member of the tetrahedral coordination of the zinc is a water molecule; this water is within hydrogen-bonding distance of the His-149 side chain. In the third monomer, His-149-Nδ is 4.0 Å away from the zinc and thus does not play a direct chelation role, but the imidazole isolates the zinc from solution and is positioned to coordinate the metal readily (Figure 2B). Point mutations of Cys-97, -99, -145, and His-149 (to alanine) show that removal of any one has a negative impact on NS3P processing (Hijikata et al, 1993; Stempniak et al., submitted). Thus, His-149 may be an integral part of zinc coordination at least during the initial folding of NS3P.

The zinc site serves to anchor the turn at βD2–βE2 (containing Cys-145 and His-149) to the interbarrel loop

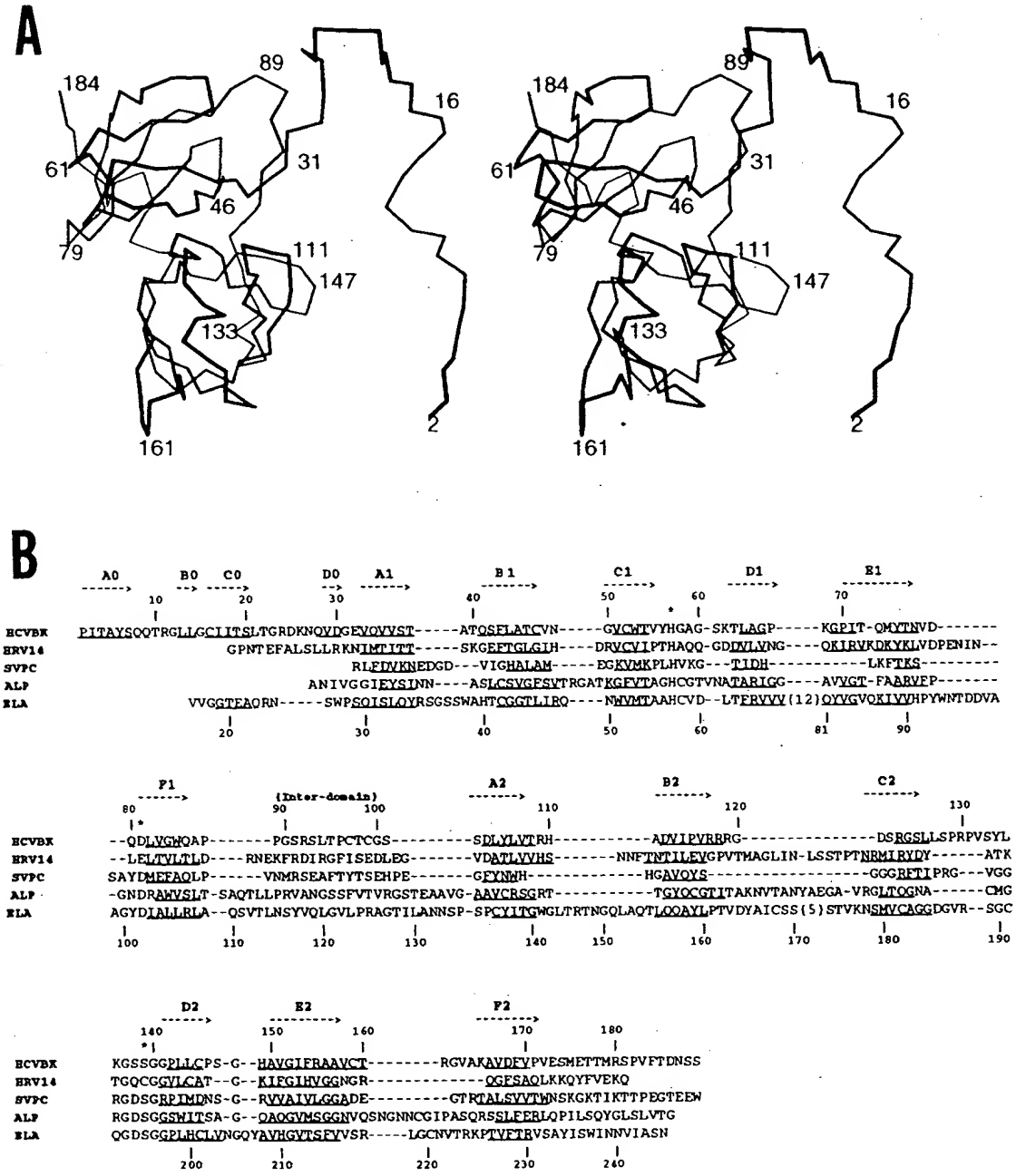


Figure 1. Trace of HCV NS3P Structure and Its Sequence Alignment with Other Proteases

(A) Side-by-side stereo view of the Cα trace for HCV NS3P (residues 2–184), drawn using Xfit (McRee, 1992). This is monomer 1 of the asymmetric unit. The trace is labeled at intervals for clarity. The N-terminal 30 residues (before the first β barrel) are extended away from the core of the protein.

(B) Alignment of HCV BK NS3P sequence with other proteinases based on superposition of crystal structures (as in Figure 3B). Residues of β strands are underlined. The general location of each β strand among all sequences is indicated by a dashed arrow; labeling of these strands follows the common convention, except for the extra N-terminal strands of HCV, which are labeled A0–D0. Catalytic triad residues are marked by an asterisk. Excess residues not shown are indicated in brackets. Abbreviations for proteins are: SVCP, Sindbis virus core protein; ALP, α-lytic proteinase; ELA, porcine elastase; HRV14, human rhinovirus type 14 3C proteinase; HCVBK, hepatitis C virus, BK strain NS3 proteinase. Amino acid numbering is for either HCVBK (top) or ELA (bottom).

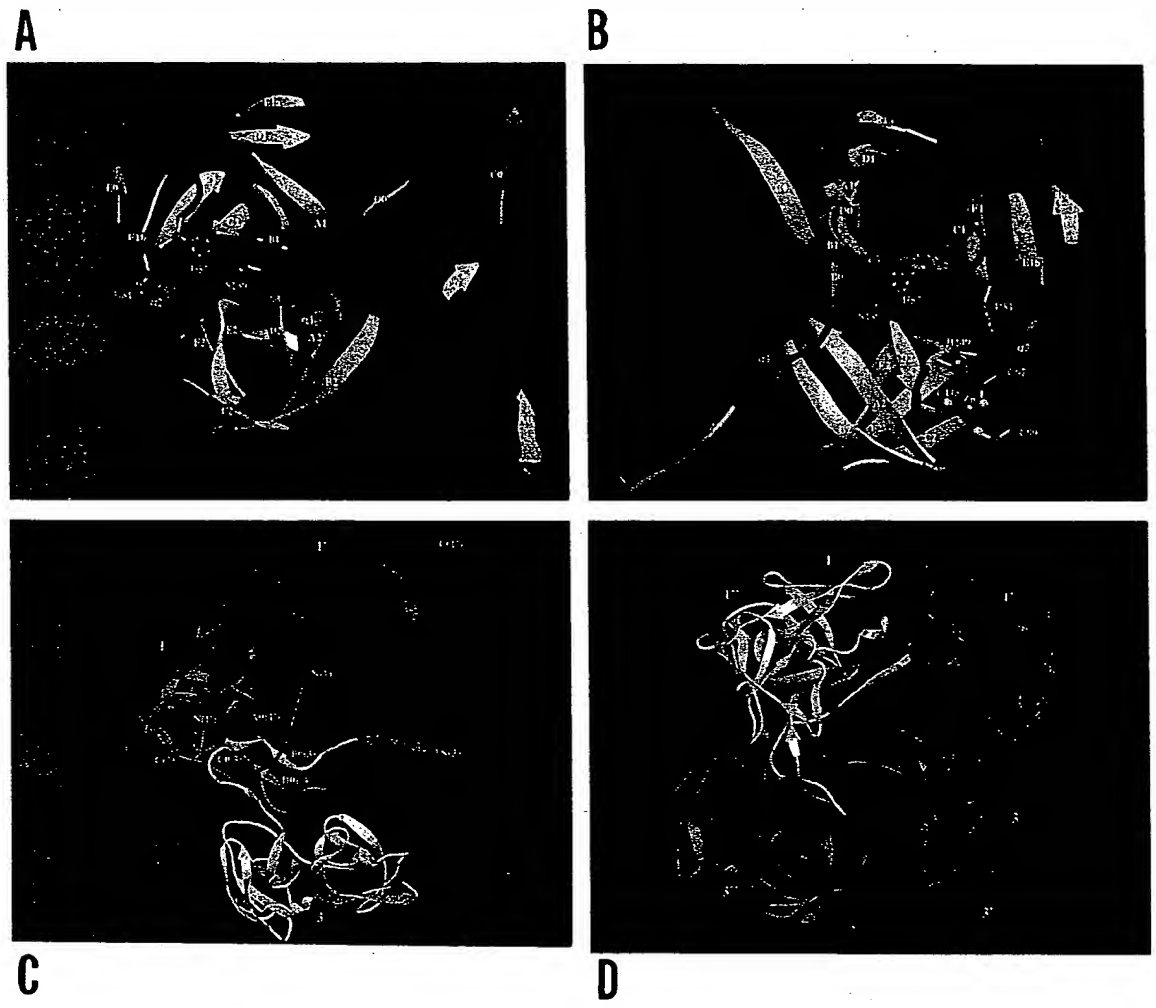


Figure 2. Overall Folding of HCV NS3P and Assembly of Molecules in the Crystal, Emphasizing N-Terminal Configuration and Exchange Displayed by Ribbons (Carson, 1991).

(A) View into the active site of monomer 1 in the asymmetric unit with catalytic triad residues labeled by the single amino acid code. Secondary structural elements are color-coded yellow for β strand, green for α helices, and blue for coil. β strands in the trypsin-like barrels (A1–F2) are labelled in the standard convention. Catalytic triad residues are shown in sphere and cylinder representation, with spheres color-coded green for carbon, blue for nitrogen, and red for oxygen.

(B) Trace of monomer 1 color-coded as in (A), rotated relative to (A) for viewing of the structural zinc-binding site. The zinc (gray sphere) is tetrahedrally coordinated by the sulfurs (yellow spheres) of cysteines 97, 99, and 145. Histidine 149, though not within bonding distance, is seen poised to complete the tetrahedral coordination (see text).

(C) Architecture of the N-terminal strand exchange. The N-terminal residues of monomer 1 (in green) extend away from the molecule permitting two antiparallel β strands, A0 of monomer 1', in blue (generated by the crystallographic 3-fold from monomer 1), and C0 of monomer 3 (in yellow), to lie on its surface. This interaction involves nonpolar side chains from the blue and yellow β strands being buried against a hydrophobic patch on the surface of the green molecule (see text). A short parallel β sheet interaction is also seen, involving D0 and B0 strands of monomers 1 and 3, respectively.

(D) Hexamer generated from monomers 1 and 3 using the crystallographic (3-fold) and noncrystallographic (2-fold) symmetry elements. In this view, the crystallographic 3-fold axis is vertical. The crystallographic trimer of monomer 1 (blue, red, and yellow) associates with the crystallographic trimer of monomer 3 (purple, light blue, and green) using N-terminal strand exchange (see text). Each molecule within the hexamer accepts two antiparallel β strands from neighboring molecules; e.g., the green molecule accepts the purple and yellow strands. The noncrystallographic 2-fold symmetry mates are red and green, yellow and purple, and blue and light blue.

(containing Cys-97 and Cys-99). Such a location for zinc, remote from the active site, implies a structural rather than a catalytic role for the metal; its effects on polyprotein processing are probably linked to accurate NS3P folding or post-folding stability of the enzyme or both. However, perturbations at the zinc site could conceivably affect the active-site conformation, because the two are linked directly through strands βD2 and βE2.

During our derivative searches (see Experimental Procedures), we found that many metals and heavy atom

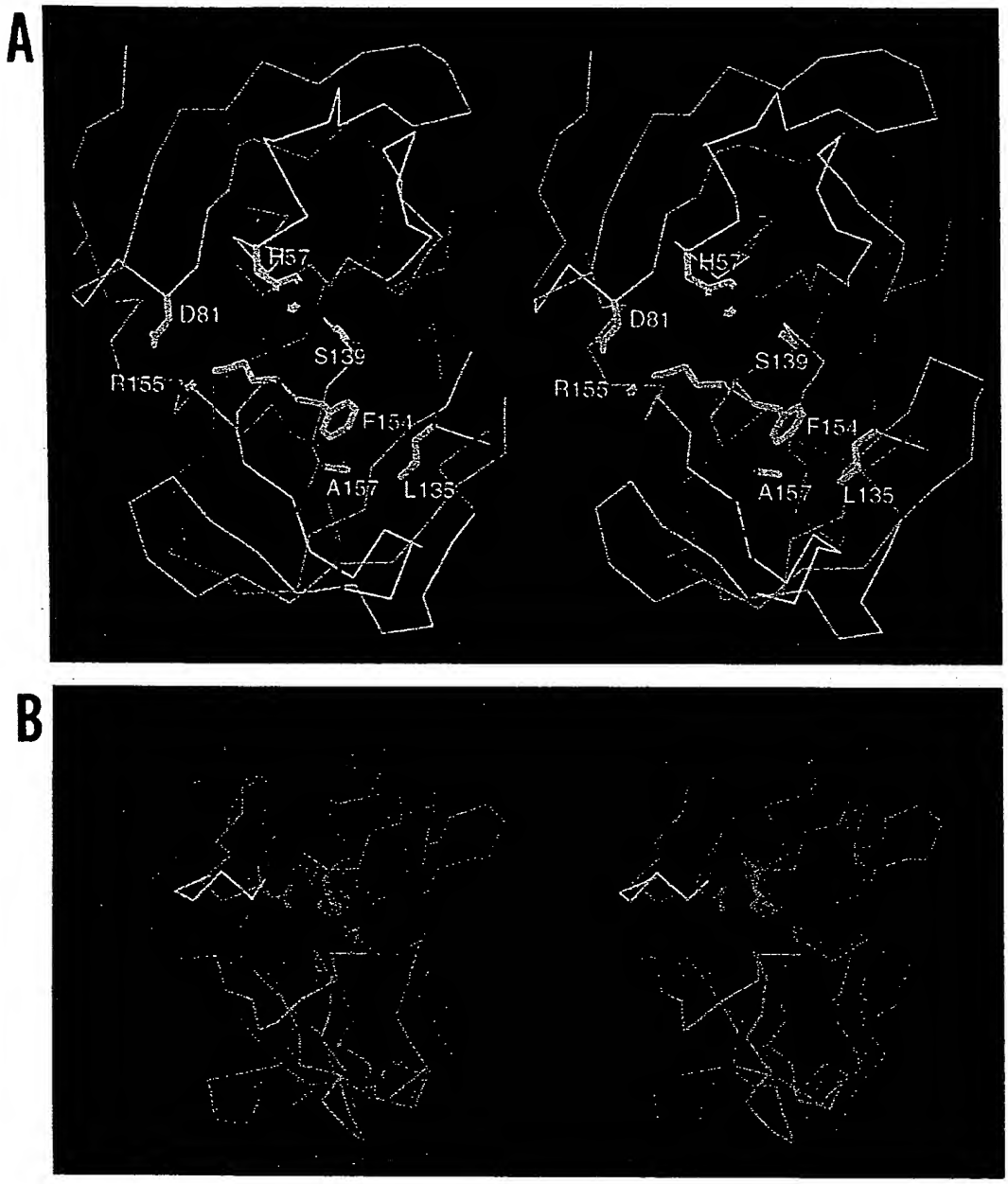


Figure 3. Active Site of NS3P and Comparison with Other Trypsin-like Serine Proteinases

(A) Side-by-side stereo view of the active site and surroundings of HCV NS3P. A Cα trace is given for clarity, with purple coloring the oxyanion loop, green for the short α helix preceeding the active-site serine, and red for the section of βE2 that interacts with substrate. Side chains, whose atoms are colored yellow (carbon), blue (nitrogen), or red (oxygen), are labeled by the single-letter amino acid code. Shown here is the active-site triad (H57, D81, S139) and residues defining the S1 specificity pocket (L135, F154, A157). Overall, this S1 pocket is relatively small and nonpolar.

(B) Superposition of HCV NS3P (green) and Sindbis core protein SCP (purple). This superposition is based on structural alignment of β structure within the barrels, which gives a root-mean-square difference on Cα's of 1.5 Å. The catalytic triad residues are shown for each molecule (thicker lines) to illustrate their similar spatial arrangement, as well as the torsional deviation of Asp-81 in NS3P. The βE2–βC2 loop (red) and βE1–βF1 loop (yellow), often determinants of P2–P5 substrate specificity in pancreatic/bacterial proteinases, are unusually short in NS3P as well as SCP. The extended N-terminus of NS3P has been omitted for clarity.

compounds bind at this zinc site in the crystal. For the case of mercury binding, we observed a disruption of the structure in this region, essentially through a disordering of the polypeptide around Cys-97–99. Other experiments showed that the zinc ion is lost when mercury binds to NS3P (presumably at cysteine) and that this binding inhibits the enzyme (Z. Hostomska, unpublished data). Copper has also been found as a potent inhibitor of NS3P (Han et al., 1995). Copper binding in our crystal occurs primarily at the zinc site, but substitution of zinc with copper is not yet confirmed. In the cases above, inhibition may result from an altered active site that in turn arises from destabilization of the zinc site during metal binding.

The zinc-coordinating amino acids of NS3P are not present in other members of Flaviviridae family such as yellow fever virus or bovine diarrhea virus but are found in the recently discovered GB viruses (Simons et al., 1995) and hepatitis G virus (Linnen et al., 1996), which are more closely related to HCV. The zinc site of NS3P may play a role analogous to the disulfide common among cellular proteinases (136–201 in elastase), which also anchors the βD2–βE2 turn to the interbarrel loop. In picornaviral 3C proteinases, which possess neither disulfides nor a structural zinc, stability in this region may be provided by the N-terminal α helix, which packs against the βD2–βE2 turn and connecting loop simultaneously.

Alignment of picornaviral 2A proteinase sequences (Yu and Lloyd, 1992; Bazan and Fletterick, 1988) has revealed two conserved motifs, CXC within the interbarrel loop and CXH at the end of βD2, which are similar in residue type and sequence position to the chelation motifs in NS3P (CXC and CXXXH). Several biochemical and biophysical studies of 2A proteinases have shown that zinc is an integral tightly bound component of the structures, required for correct folding and stability (but not involved in catalysis) and probably chelated by residues of the two motifs (Yu and Lloyd, 1992; Sommergruber et al., 1994; Voss et al., 1995). Thus, in the picornaviral 2A proteinases there is probably a zinc site with general location, coordination geometry, and stabilizing function that parallels the NS3P case.

Active Site

In the crystal structure of NS3P, the catalytic triad residues (His-57, Asp-81, Ser-139), the "oxyanion-stabilizing loop" (135–139), and strand βE2 forming one side of the specificity pocket, together have the same relative spatial positions as in other trypsin-like proteinases (Figure 3). Mutation experiments have identified the presumed NS3P triad residues as essential to proteolytic activity (Hijikata et al., 1993).

The imidazole of His-57, expected to extract a proton from nucleophile Ser-139 during the enzymatic reaction, is oriented toward Ser-139 but is not close enough to form the hydrogen bond often observed in proteinase structures. The stretch of residues from 57 to 63 has greater mobility than most loops in the structure, as indicated by higher temperature factors; however, the conformation we have built here is consistent with our experimental density maps and resembles the helical turn found in other proteinases. Flexibility in this region may be related to a lack of structural anchoring compared with cellular proteinases, where a conserved disulfide links Cys-58 to Cys-42 on strand βB1.

The side chain of Asp-81, expected to provide charge stabilization for His-57 after deprotonation of Ser-139, is oriented away from His-57 in the crystal structure of NS3P. However, the Ser-139-Cα to Asp-81-Cα distance of 11.4 Å is close to that in rhinovirus 3C proteinase (11.4 Å) and also Sindbis core protein (10.9 Å), while only slightly longer than that in cellular proteinases (approximately 10 Å). Asp-81 forms an ion pair with Arg-155 (on βE2), the latter being a conserved amino acid among HCV sequences. Arg-155 in NS3P corresponds to conserved Ser-214 of other proteinases, where the hydroxyl of the serine hydrogen-bonds to the carboxylate of triad member Asp-102. Interestingly, the rotation of Asp-81 away from His-57 observed in NS3P, and the Asp-81–Arg-155 interaction, are paralleled in the crystal structure of 3C proteinase from hepatitis A (Allaire et al., 1994), where the aspartate points away from histidine while interacting with a lysine on strand βF2. In the hepatitis A case, however, it was suggested that the aspartic acid is not crucial for histidine charge stabilization because the cysteine nucleophile is easier to deprotonate than serine. In NS3P, only minor readjustments of amino acids 80–82 would position the Asp-81 carboxylate closer to the His-57 side chain, as probably required for catalysis. We propose that a classic catalytic triad configuration for His-57/Asp-81/Ser-139 will exist during substrate cleavage by NS3P in vivo and in vitro and that the positional deviation of Asp-81 we observe is a consequence of an apo-enzyme structure or the crystallization conditions or both.

The sequence around Ser-139 in NS3P (Gly-Ser-Ser-Gly-Gly) follows the conserved GXSGG motif seen in trypsin-like serine (and cysteine) proteinases, and the polypeptide backbone conformation of the oxyanion-stabilizing loop (135–139) is nearly identical to that in other proteinases (see Figures 3B and 6). Alignment of all backbone atoms of NS3P residues 135–139 with the corresponding atoms of elastase (191–195) gives a root-mean-square difference of 0.71 Å. The βC2–βD2 loop, always variable among proteinases, is unique in NS3P because it contains one turn of α helix (residues 131–134) just prior to the oxyanion loop. This helix could add conformational stability at the binding site, perhaps analogous to the conserved disulfide found in many proteinases (involving Cys-191) that anchors the oxyanion loop to the βE2–βF2 turn.

The cleavage performed by NS3P is between polyprotein substrate residues Cys/Ser (at 4B/5A and 5A/5B sites), Cys/Ala (at 4A/4B), or Thr/Ser (at 3/4A). Our structure shows a specificity pocket that is shallow and nonpolar, formed primarily by the side chains of invariant residues Phe-154, Ala-157, and Leu-135 (Figure 3A). Homology modeling had predicted this type of S1 pocket, involving at least Phe-154 (Pizzi et al., 1994; see also "Crystallographic Phasing Strategies," below). Phe-154 corresponds to position 213 of the cellular proteinases, where it is typically a small hydrophobic amino acid. Picornavirus 3C proteinases have histidine at 213; thus, they show a closer steric resemblance to NS3P at this

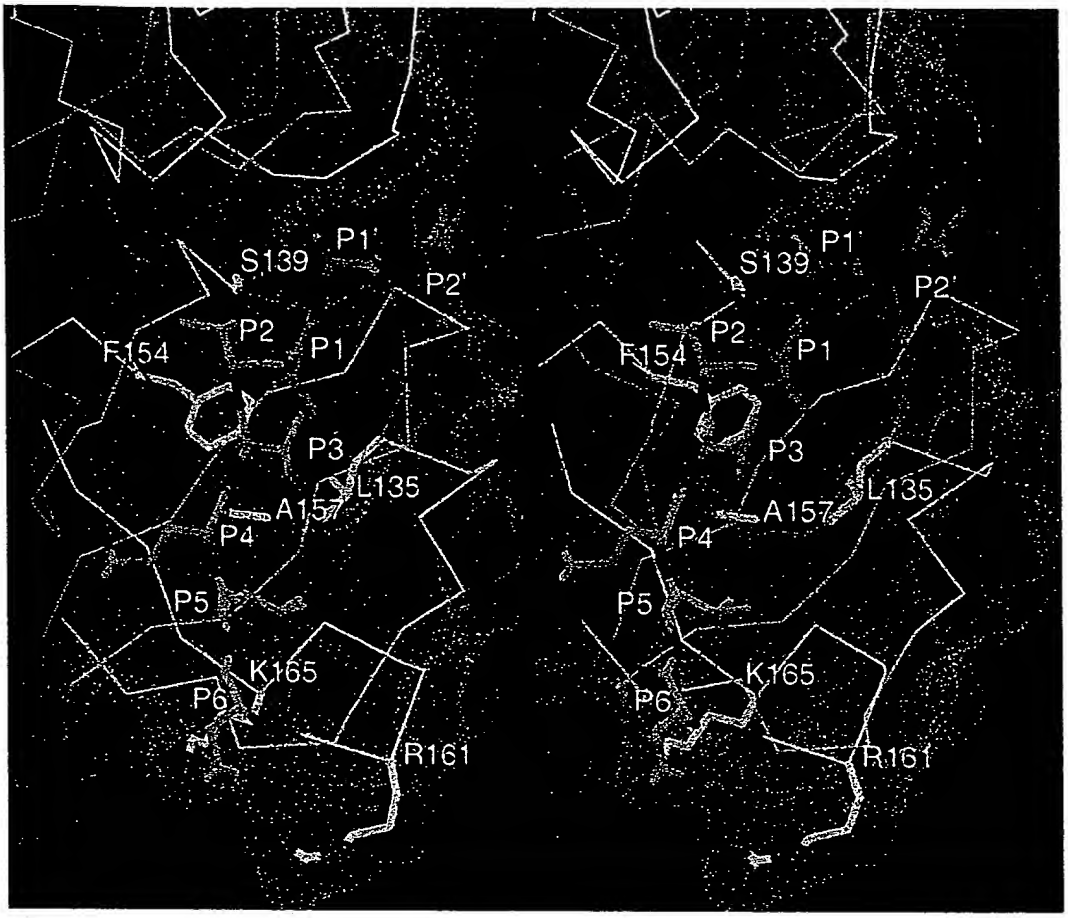


Figure 4. Model of a Bound Polypeptide Substrate Containing the NS3/4A Junction

Side-by-side stereo view of the NS3/4A substrate (color-coded: green, C; red, O; blue, N) modeled into the active site of HCV NS3P, which is covered by a molecular surface in white dots (generated by the MS program [Connolly, 1983] using a 1.6 Å probe radius). The Cα trace of NS3P is shown in yellow, with the oxyanion-stabilizing loop in purple and the extended strand βE2 in red. NS3P side chains (color-coded: yellow, C; red, O; blue, N) are shown for residues comprising the S1 specificity pocket (L135, F154, A157), the potential P6 recognition elements (R161, K165), and active-site serine (S139). The substrate is labeled according to the standard P/P' convention, while the HCV residues are labeled in the single-letter amino acid code.

position. The Leu-135 side chain forms the bulk of the remaining S1 pocket and is located approximately at position 190 of cellular proteinases, often a small polar amino acid and specificity determinant. Ala-157 and Val-167 of NS3P correspond to positions 216 and 226, respectively, of cellular proteinases, which are usually glycine to accommodate long P1 side chains; elastase, which recognizes a small nonpolar P1 side chain and has Val-216/Thr-226, most closely parallels NS3P at these positions.

Substrate Modeling in NS3P

Hydrolysis of the NS3/4A peptide bond is the first (in *cis*) cleavage event by NS3P, involving a threonine at the P1 position of the substrate; for subsequent cleavage events, the P1 residue is cysteine. A model of the Michaelis complex between a polypeptide substrate containing the NS3/4A cleavage site and NS3P was constructed (Figure 4), based on the known binding mode of polypeptide inhibitors of trypsin-like enzymes (Read and James, 1986). This model suggests the possibility of a favorable interaction between the OH (or SH) of the P1 side chain and the electron-rich π clouds on the aromatic ring of nearby Phe-154. A sulfhydryl–aromatic interaction was earlier proposed as a possible NS3P substrate recognition mechanism (Pizzi et al., 1994). Interaction between a threonine hydroxyl and a tyrosine aromatic ring has been observed (Ji et al., 1992; Liu et al., 1993). Numerous observations and calculations support the notion that hydrogen bonding between proton donors and aromatic rings plays a significant role in protein stability (Burley and Petsko, 1986; Levitt and Perutz, 1988).

Strand βE2 of NS3P, expected to form β sheet antiparallel hydrogen bonds with the substrate over P2–P3, is more extended than in most proteinases. Its length most closely resembles the corresponding β strand in Sindbis core protein. We hypothesize that the NS3P enzyme–

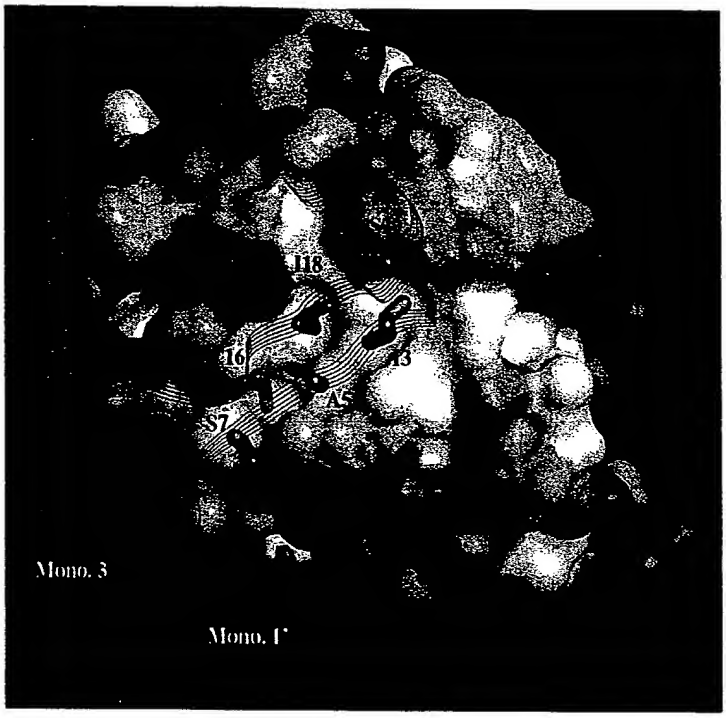


Figure 5. Solvent-Accessible Surface of NS3P Monomer 1, Colored by Hydrophobicity, Such That Nonpolar Residues Are White, Charged Residues Are Deep Magenta, and Polar Residues Are Medium Shades of Magenta

The central white region is the hydrophobic patch (approximately 400 Å^2) discussed in the text. In the crystal, there are two N-terminal strands from neighboring molecules bound to this patch, namely βC0 from monomer 3 (blue ribbon) and βA0 from monomer 1' (green ribbon). Strand βC0 buries Cys-16 and Ile-18, while strand βA0 buries Ile-3 and Ala-5, into the center of the hydrophobic surface patch. Residues Ser-20 of βC0 and Ser-7 of βA0 lie at the edge of the interface and hydrogen-bond to monomer 1. Also shown in yellow (at bottom horizon of surface) is the P3' residue of a modeled substrate polypeptide (see Figure 4), included here as a directional reference to the active site.

substrate β-interaction at P2–P3 continues for another three residues to P6 (Figure 4), for two reasons. First, the consensus acidic residue at substrate P6 could interact with the Arg-161 and Lys-165 of NS3P, which are relatively isolated on the extended βE2–βF2 turn. Secondly, a continuous P2–P6 main-chain β-interaction might compensate for the apparent lack of P2–P5 side chain to enzyme interactions, which results from the relative shortness of the βB2–βC2 and βE1–βF1 loops in NS3P.

N-Terminal Configuration

A quite unexpected feature of the crystal structure of NS3P is the configuration of its long N-terminus (the first 30 amino acids), which extends away from the protein and contains β strands that interact with neighboring molecules (see Figure 2). Monomers 1 and 3 of the asymmetric unit ultimately assemble into a hexamer via a combination of crystallographic symmetry (a trimer is generated from either monomer) and noncrystallographic symmetry (association of these two trimers). The order of this assembly is unknown. A similar hexamer is formed from six copies of monomer 2 by (32) crystallographic symmetry operators. However, weaker electron density at the N-termini in this hexamer precludes the building of polypeptide chains between copies of monomer 2.

One consequence of the strand exchange is that a hydrophobic patch on the surface of each molecule is covered by two relatively nonpolar β strands from N-termini of a different neighbor (Figures 2C and 5). Specifically, monomer 3 contributes βC0 (residues 16–20; sequence CIITS) to the patch on monomer 1, with Cys-16 and Ile-18 buried and Ser-20 partially buried at the interface boundary. Antiparallel to this strand, and also lying against the patch, is βA0 (residues 2–7; sequence PITAYS) from a crystallographically related copy of monomer 1, which buries Ile-3 and Ala-5 into the patch while partially burying Ser-7. An additional association, apart from the patch, involves a short two-stranded parallel β-interaction (βB0: βD0 between monomers 1 and 3) that occurs in between trimers, where two traversing N-termini approach one another (Figure 2C).

The "hydrophobic patch" is a long narrow (approximately 20 Å × 8 Å) shallow valley with a continuous nonpolar molecular surface area of about 400 Å^2, located mainly on the second β barrel, and composed of residues primarily from strands βA2, βB2, and βD2 but also two residues from βA1 (Figure 5). Specifically, the patch is formed by the side chains of conserved residues Val-33, Val-35, Leu-44, Leu-94, Val-107, Leu-127, Ala-111, Val-113, Pro-115, Pro-142, Leu-144, Tyr-105 (ring portion), and Arg-109 (alkyl chain portion).

Since hydrophobic interactions are often key components of molecular recognition, the association of the two β strands with the hydrophobic patch on NS3P may mimic one or more molecular interactions that occur during HCV polyprotein processing, rather than being an artifact of crystallization. The patch could qualify as a functionally important binding site because it is defined by residues from strand βD2, which contains the nucleophile, and by the antiparallel strands βA2/βB2, which together directly support and hydrogen-bond to the oxyanion loop. It is known that the NS4A domain (immediately downstream of NS3) binds to NS3P during processing, and this facilitates the cleavages by NS3P. Furthermore, the N-terminus of NS3P itself participates in the cofactor modulation of NS4A (Koch et al., 1996).

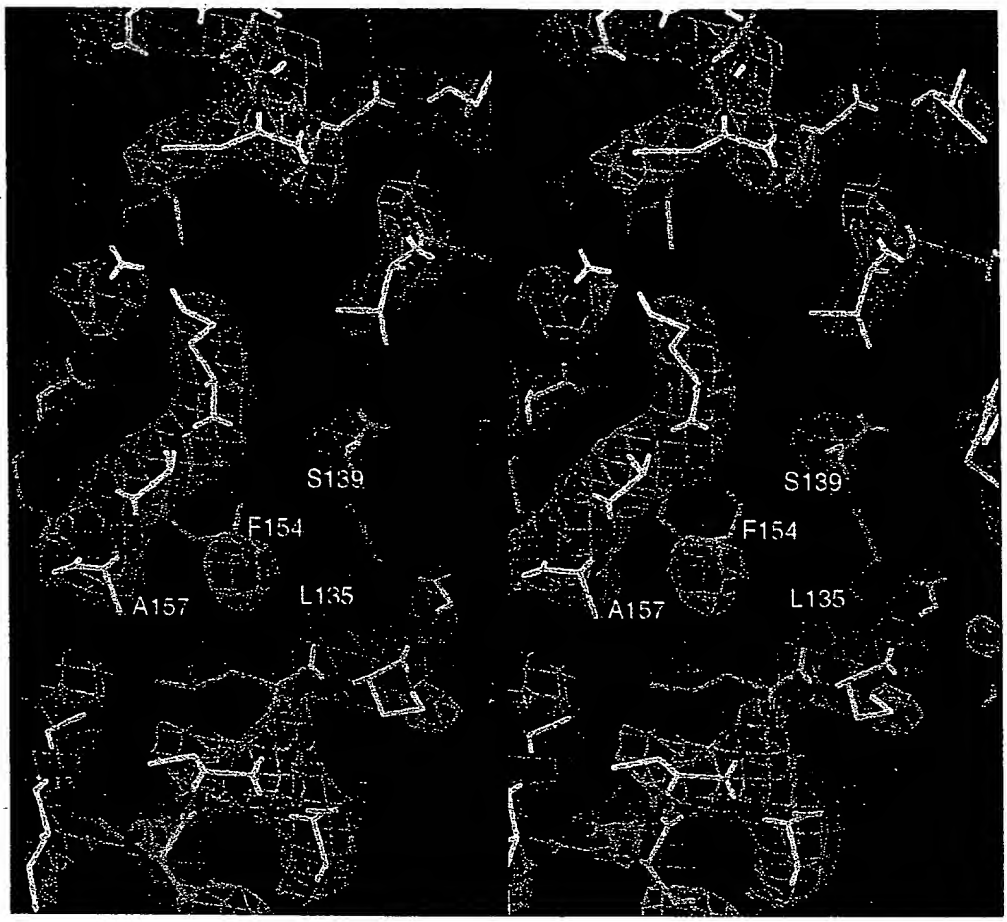


Figure 6. Section of the 2.7 Å Experimental Electron Density Map

Side-by-side stereo view of NS3P (color-coded: yellow, C; red, O; blue, N) in the vicinity of the active site, superimposed on the 2.7 Å experimental ISIRAS map (purple mesh), contoured at 1.5 σ. A clear definition of main chain, side chain, and, frequently, carbonyl direction exists throughout the majority of the map.

Therefore, one possibility is that this hydrophobic patch forms part of a site through which the N-terminus of NS3P or the cofactor NS4A or both modulate the activity of NS3P.

Experimental Procedures

Protein Expression and Purification

HCV BK NS3 proteinase (residues 1–189) was expressed in E. coli as a soluble protein. The isolation of a pure protein involved three chromatographic steps: Fast Flow SP Sepharose; FPLC Mono-S, and Sepharose S200. The final preparation of HCV BK 1–189 migrated as a single entity on SDS–polyacrylamide gel electrophoresis in reducing and nonreducing buffer system, with an apparent molecular mass of 20,000 kDa. The specific activity of the pure sample of HCV BK 1–189 was about 50 nmol/min/mg using a 5A/5B 15 residue synthetic peptide substrate (C. Lewis, unpublished data). Automated Edman degradation gave an N-terminal sequence PITAYS; thus, the initiation methionine and alanine residues from the original sequence are missing. Details of protein purification and enzyme characterization will be published elsewhere.

Protein Crystallization and X-Ray Data Collection

Crystals were grown at 4°C by the hanging-drop vapor diffusion method, using plastic Linbro tissue culture plates. Aliquots (5 μl) of 10 mg/ml of protein HCV BK 1–189 in 50 mM sodium acetate, 10 mM dithiothreitol, 350 mM NaCl (pH 6.0) were mixed with 5 μl of reservoir containing 5% PEG 400, 3.5 M NaCl, 150 mM Tris–HCl (pH 8.5). Crystals appeared after 2–3 weeks, and they grew as hexagonal rods with typical dimensions 0.1 ×,0.1 × 0.6 mm.

The crystals belong to space group R32 with hexagonal cell parameters of a = b = 133 Å, c = 223 Å. There are three molecules per asymmetric unit, which gives a calculated solvent content of 62%. The extreme sensitivity of heavy atom–soaked crystals to X-rays, plus the eventual use of synchrotron radiation, prompted a uniform application of cryo-cooling techniques (−170°C). In-house data were obtained from a 30 cm MAR imaging plate and processed with DENZO and SCALEPACK (Otwinowski and Minor, 1993), and complete data to 3.5 Å (derivatives) or 3.0 Å (native) were routinely obtained. Synchrotron sources at Photon Factory, Japan, and European Synchrotron Radiation Facility, France, were ultimately used to obtain higher resolution native and derivative data (see Table 1).

Crystallographic Phasing Strategies

During heavy-atom screening in house, most trials led to isomorphous difference Patterson maps containing peaks at similar positions; these peaks were also seen in several mercury and gold anomalous difference Pattersons. Compounds with a wide variety of binding mechanisms led to the same peaks. Nonisomorphism was a significant problem, and data from short soaks frequently

Table 1. Crystallographic Data Statistics

Parameters	$HgCl_2$/Native (1)	BMDA[a]	Native (2)
	Diffraction data		
Heavy atom (mM)	0.05	6.0	
Soaking time (hrs)	25.0	41.0	
Synchrotron facility[b]	Photon Factory	Photon Factory	ESRF
Data collection device[c]	Fuji ip	Fuji ip	Mar ip
Wavelength (Å)	1.0000	1.0064	0.995
Maximum resolution (Å)	2.7	2.5	2.4
Data completeness (%)	93.8	93.9	97.8
I/σI at resolution limit	3.8	4.4	3.9
Observations (No.)	93,997	135,383	96,285
Unique reflections	19,945	24,998	29,437
Rsym[d] (%)	7.5	7.7	7.9
Riso[e] (%)		16.5	
	SIRAS statistics to 2.7 Å		
R_{cullis}[f]		0.56	
R_{cullis}[g] anomalous		0.32	
FOM[h]		0.69	
solvent flattened FOM		0.87	
rms f/Eiso		2.34	
rms f''/Eano		2.86	
Sites (No.)		9	

[a] 3,6-bis(mercurimethyl) dioxane acetate.
[b] Beamline 18B, Photon Factory Tsukuba City, Japan; Beamline 4, ESRF European synchrotron radiation facility in Grenoble, France.
[c] Fuji ip: data measured using a Weissenberg camera in oscillation mode, Mar ip: Mar research image plate system. Data from both systems were processed with DENZO and SCALEPACK (Otwinowski and Minor, 1996).
[d] R_{sym} is the unweighted R factor on I between symmetry mates.
[e] $R_{iso} = \Sigma(|F_{der}|-|F_{nat}|)/\Sigma|F_{nat}|$ where F_{der} and F_{nat} are derivative and native structure factor amplitudes, respectively.
[f] $R_{cullis} = \Sigma||F_{der} +/- F_{nat}|-F_{Hcalc}||/\Sigma|F_{der} +/- F_{nat}|$ for all centric reflections.
[g] R_{cullis} anomalous was calculated for the top 25 percent largest Bijvoet differences.
[h] FOM is the figure of merit.

had to serve as "native" relative to longer soaks as "derivative." A total of three major sites per asymmetric unit was eventually deduced from Patterson maps using HASSP (Terwilliger et al., 1987). Later, it was found that each site lay near a cluster of cysteines that formed a structural zinc site and that this region was disrupted by heavy-atom binding. The best in-house phases were obtained by a combination of isomorphous and anomalous data from $HgCl_2$ soaks. Heavy-atom site refinement using PHASES (Furey and Swaminathan, 1990) at 5 Å gave an overall figure of merit, phasing power isomorphous/anomalous, and centric R factor of 0.75, 2.5/3.0, and 0.45, respectively. This was used to initiate solvent flattening (Wang, 1985) and to check the handedness of the sites. Phasing information beyond 5 Å was poor, but the 5 Å electron density maps showed well-defined and separated molecular envelopes, indicating three molecules per asymmetric unit. The heavy-atom positions were used to determine noncrystallographic operators, so that 3-fold averaging combined with phase extension to 3.5 Å could be initiated using PHASES. While such maps could not be used for tracing, they did show features that suggested a trypsin-like structure, namely two globular domains, one of which resembled a thick-walled barrel. This permitted a positioning of our NS3P homology model into the unit cell. This model was based on rhinovirus 3C proteinase and generated with LOOK (Molecular Applications Group, Stanford, CA). When higher resolution maps became available (see below), it was found that this model had been correctly positioned and could serve as a guide for building the structure.

Data collected at Photon Factory from two mercury-soaked crystals, using 3,6-bis(mercurimethyl) dioxane (BMDA) and $HgCl_2$, provided the higher resolution phases necessary to produce traceable maps. The isomorphous signal from differences between BMDA and $HgCl_2$ ($HgCl_2$ being more native-like), combined with the anomalous signal from the BMDA data, led to the best results (Table 1). For BMDA, a total of six more sites per asymmetric unit was found by difference Fourier methods; these were spatially close to the first set of three. Phasing statistics indicated usable experimental phases to 2.7 Å. Density maps calculated to 3.0 Å and 2.7 Å revealed a nearly complete trace for the protein, with abundant side-chain information, and confirmed that NS3P folds with a double-β barrel trypsin-like motif.

Model Building and Refinement

Based on 3 Å and 2.7 Å solvent-flattened electron density maps, a model was built for one molecule of the asymmetric unit into the region with best connectivity, using FRODO (Jones, 1978) and XFIT (McRee, 1992). This model was then copied to the other two positions via the previously refined noncrystallographic symmetry (NCS) operators and adjusted locally where necessary. At this point, each monomer consisted of a trypsin-like core with two disconnected parallel elongated strands lying upon one surface, which represented in some manner the N-terminus (approximately 29 residues). Side-chain density in these strands clearly defined a unique sequence fit and direction, which was inconsistent with the strands belonging to the same (nearest) molecule. For example, density for an approximately eight-amino acid β strand ending close to residue 30 (the start of the first barrel) could not be fit with the sequence 21–29 as expected; instead, it was fit clearly by 14–21 (most notably at Cys-16) and running in the opposite direction. This situation, along with nearly continuous density extending between molecules (before 14 and after 21), led to the N-terminal strand-exchange structure of NS3P.

Initial crystallographic refinement with XPLOR (Brünger, 1992b) and manual readjustment of the model were done at 2.7 Å using the $HgCl_2$ data (this being more native-like), with NCS restraints applied to the three monomers of the trimer. Later, native data to 2.4 Å resolution from European Synchrotron Radiation Facility was used, and the three molecules were refined with only weak NCS restraints, the weight chosen to optimize the free R value (Brünger, 1992a). XPLOR treatment included simulated annealing ("slow-cooling" from progressively lower temperatures as the model improved), positional refinement, and then restrained individual temperature

factor refinement. Rebuilding was done from annealed-omit maps and (2Fo-Fc) maps, with continual reference to the high quality experimental maps. With the European Synchrotron Radiation Facility native data, it was possible to construct a zinc site, based on 6 σ density in (Fo-Fc) maps, between two neighboring surface loops where cysteines 97, 99, and 145 were known to be located. The zinc site was eventually refined with only nonbonded parameters (no explicit chelation bonds) and remained stable as three sulfhydryls surrounding the zinc, with sulfur–zinc distances of 2.0–2.5 Å. Verification of the N-terminal trace and strand exchange was made for two molecules out of three in the asymmetric unit (our monomers "1" and "3"). For monomer "1," the only disordered residues are the last five (out of 188); this is the molecule chosen for the Discussion. In the other two molecules, density beyond residue 177 is poorly defined. Using only residues in the β barrel motifs of each molecule, the root-mean-square diffence in Cα positions between monomers ranges from 0.5 Å for monomers 1–3 to 0.7 Å for monomers 1–2.

A total of 125 water molecules and seven chloride ions were eventually included during the refinement process of the trimer (5058 protein atoms). The chlorides were chosen to replace those waters whose B factors had refined very low and which had locations near electrostatically unpaired basic side chains. The R factor is 22.5% over 8–2.4 Å (European Synchrotron Radiation Facility data; 25,000 reflections with $I > 2.0\ \sigma[I]$), and the free R factor is 32% (2,000 reflections). Bond and angle deviations are 0.02 Å and 2.1°, respectively, as determined by XPLOR using the Engh and Huber (1991) parameters. PROCHECK (Laskowski et al., 1993) analysis of ϕ/ψ angles indicates no residues in disallowed regions and two in "generously allowed" regions. The average temperature factor is 17 Å^2 for main-chain atoms and 20 Å^2 for side-chain atoms.

Acknowledgments

Correspondence should be addressed to Z. Hostomska. We wish to thank the following individuals: Drs. D. Knighton and C. Kissinger for help with data collection and analysis; M. Stempniak, B. Nodes, N. Tajima, S. Rahmati, B. Aust, and G. Hudson for excellent technical assistance; Drs. S. Reich, E. Villafranca, P. Dragovich, and C. Lewis for valuable discussions; Dr. D. Matthews for critical reading of the manuscript; and Drs. H. Tsuge, M. Miyano, H. Ago, N. Watanabe, M. Suzuki, N. Sakabe, and E. Inagaki for their assistance with data collection at Photon Factory. This study was in part supported by the Sakabe project of Tsukuba Advanced Research Alliance.

Received September 13, 1996; revised September 24, 1996.

References

Allaire, M., Chernaia, M.M., Malcolm, B.A., and James, M.N.G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature *369*, 72–76.

Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1994). Kinetic and structural analyses of hepatitis C virus polyprotein processing. J. Virol. *68*, 5045–5055.

Bartenschlager, R., Ahlborn-Laake, L., Yasargil, K., Mous, J., and Jacobsen, H. (1995). Substrate determinants for cleavage in *cis* and in *trans* by the hepatitis C virus NS3 proteinase. J. Virol. *69*, 198–205.

Bazan, J.F., and Fletterick, R.J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc. Natl. Acad. Sci. USA *85*, 7872–7876.

Brünger, A. (1992a). Free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Brünger, A. (1992b). XPLOR Version 3.1: A System for X-Ray Crystallography and NMR. (New Haven, Connecticut: Yale University).

Burley, S.K., and Petsko, G.A. (1986). Amino–aromatic interactions in proteins. FEBS Lett. *203*, 139–143.

Carson, M. (1991). Ribbons 2.0. J. Appl. Cryst. *24*, 958–961.

Choi, H.-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossman, M.G., and Wengler, G. (1991). Structure of Sindbis virus coat protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature *354*, 37–43.

Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M. (1989). Isolation of a cDNA clone derived from blood-borne non-A, non-B viral hepatitis. Science *244*, 359–362.

Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G., and Houghton, M. (1991). Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA *88*, 2451–2455.

Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. Science *221*, 709–713.

D'Souza, E.D.A., Grace, K., Sangar, D.V., Rowlands, D.J., and Clarke, B.E. (1995). *In vitro* cleavage of hepatitis C virus polyprotein substrates by purified recombinant NS3 protease. J. Gen. Virol. *76*, 1729–1736.

Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Cryst. *A47*, 392–400.

Failla, C., Tomei, L., and De Francesco, R. (1994). Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. J. Virol. *68*, 3753–3760.

Fujinaga, M., Delbaere, T.J., Brayer, G.D., and James, M.N.G. (1985). Refined structure of α-lytic protease at 1.7 Å resolution: analysis of hydrogen bonding and solvent structure. J. Mol. Biol. *183*, 479–502.

Furey, W., and Swaminathan, S. (1990). "PHASES"—a program package for the processing and analysis of diffraction data from macromolecules. ACA Meeting Summaries *73* (New York: American Crystallographic Association).

Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M., and Rice, C.M. (1993). Characterization of the hepatitis C virus–encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J. Virol. *67*, 2832–2843.

Han, D.S., Hahm, B., Rho, H.-M., and Jang, S.K. (1995). Identification of the protease domain in NS3 of hepatitis C virus. J. Gen. Virol. *76*, 985–993.

Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K., and Shimotohno, K. (1993). Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J. Virol. *67*, 4665–4675.

Houghton, M. (1996). Hepatitis C viruses. In Fields Virology, Third Edition, B.N. Fields, D.M. Knipe, and P.M. Howley, eds. (New York: Raven Press), pp. 1035–1058.

Ji, X., Zhang, P., Armstrong, R.N., and Gilliland, G.L. (1992). The three-dimensional structure of a glutathione S-transferase from the Mu class: structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2 Å resolution. Biochemistry *31*, 10169–10184.

Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. J. Appl. Cryst. *11*, 268–272.

Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K. (1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA *87*, 9524–9528.

Koch, J.O., Lohman, V., Herian, U., and Bartenschlager, R. (1996). *In vitro* studies on the activation of the hepatitis C virus NS3 proteinase by the NS4A cofactor. Virology *221*, 54–66.

Laskowski, R.J., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl. Cryst. *26*, 283–291.

Levitt, M., and Perutz, M.F. (1988). Aromatic rings as hydrogen bond acceptors. J. Mol. Biol. *201*, 751–754.

Lin, C. and Rice, C.R. (1995). The hepatitis C virus NS3 serine proteinase and NS4A cofactor: establishment of a cell-free trans-processing assay. Proc. Natl. Acad. Sci. USA *92*, 7622–7626.

Lin, C., Pragai, B.M., Grakoui, A., Xu, J., and Rice, C.M. (1994). Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics. J. Virol. *68*, 8147–8157.

Linnen, J., Wages, J., Jr., Zhang-Keck, Z.-Y., Fry, K.E., Krawczynski,

K.Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J.W.-K., Young, L., Piatak, M., Jr., Hoover, C., Fernandez, J., Chen, S., Zou, J.-C., Morris, T., Hyams, K.C., Ismay, S., Lifson, J.D., Hess, G., Foung, S.K.H., Thomas, H., Bradley, D., Margolis, H., and Kim, J.P. (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science *271*, 505–508.

Liu, S., Ji, X., Gilliland, G.L., Stevens, W.J., and Armstrong, R.N. (1993). Second-sphere electrostatic effects in the active site of glutathione S-transferase: observation of an on-face hydrogen bond between the side chain of threonine 13 and the π-cloud of tyrosine 6 and its influence on catalysis. J. Am. Chem. Soc. *115*, 7910–7911.

Matthews, D.A., Smith, W.W., Ferre, R.A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J.E., Janson, C.A., McElroy, H.E., Gribskov, C.L., and Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell *77*, 761–771.

McRee, D.E. (1992). XtalView: a visual protein crystallographic system for X11/Xview. J. Mol. Graph. *10*, 44–47.

Meyer, E., Cole, G., Radhakrishnan, R., and Pepp, O. (1988). Structure of native porcine pancreatic elastase at 1.65 Å resolution. Acta Cryst. *B44*, 26–38.

Miller, R.H., and Purcell, R.H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA *87*, 2057–2061.

Otwinowski, Z., and Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. *276*, 307–326.

Pizzi, E., Tramontano, A., Tomei, L., La Monica, N., Failla, C., Sardana, M., Wood, T., and De Francesco, R. (1994). Molecular model of the specificity pocket of the hepatitis C virus protease: implications for substrate recognition. Proc. Natl. Acad. Sci. USA *91*, 888–892.

Read, R.J., and James, M.N.G. (1986). Introduction to the protein inhibitors: X-ray crystallography. In Proteinase Inhibitors, A.J. Barret and G. Salvesen, eds. (Amsterdam; New York; Oxford: Elsevier Science Publishers BV), pp. 301–336.

Schwabe, J.W.R., and Klug, A. (1994). Zinc mining for protein domains. Struct. Biol. *1*, 345–349.

Simons, J.N., Leary, T.P., Dawson, G.J., Pilot-Matias, T.J., Muerhoff, A.S., Schlauder, G.G., Desai, S.M., and Mushahwar, I.K. (1995). Isolation of novel virus-like sequences associated with human hepatitis. Nature Med. *1*, 564–569.

Simmonds, P., Smith, D.B., McOmish, F., Yap, P.L., Kolberg, J., Urdea, M.S., and Holmes, E.C. (1994). Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS5 regions. J. Gen. Virol. *75*, 1053–1061.

Sommergruber, W., Casari, G., Fessl, F., Seipelt, J., and Skern, T. (1994). The 2A proteinase of human rhinovirus is a zinc containing enzyme. Virology *204*, 815–818.

Steinkühler, C., Tomei, L., and De Francesco, R. (1996). *In vitro* activity of hepatitis C virus protease NS3 purified from recombinant baculovirus-infected SF9 cells. J. Biol. Chem. *271*, 6367–6373.

Suzich, J.A., Tamura, J.K., Palmer-Hill, F., Warrener, P., Grakoui, A., Rice, C.M., Feinstone, S.M., and Collett, M.S. (1993). Hepatitis C virus NS3 protein polynucleotide–stimulated nucleoside triphosphatase and comparison with related pestivirus and flavivirus enzymes. J. Virol. *67*, 6152–6158.

Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I., and Okayama, H. (1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. *65*, 1105–1113.

Terwilliger, T.C., Kim, S.-H., and Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. Acta Cryst. *A43*, 1–5.

Vallee, B.L., and Auld, D.S. (1990). Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry *29*, 5647–5659.

van Doorn, L.J. (1994). Review: molecular biology of the hepatitis C virus. J. Med. Virol. *43*, 345–356.

Voss, T., Meyer, R., and Sommergruber, W. (1995). Spectroscopic characterization of rhinoviral protease 2A: Zn is essential for the structural integrity. Protein Sci. *4*, 2526–2531.

Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90–112.

Yu, S.F., and Lloyd, R.E. (1992). Characterization of the roles of conserved cysteine and histidine residues in poliovirus 2A protease. Virology *186*, 725–735.

The EMBO Journal vol.9 no.8 pp.2 . 2638, 1990

Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing *in vivo*

Raoul J.de Groot, W.Reef Hardy, Yukio Shirako[1] and James H.Strauss[2]

Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA

[1]Present address: Center for Agricultural Molecular Biology, Cook College, Rutgers University, New Brunswick, New Jersey 08903

Communicated by A.van Kammen

[2]Corresponding author

The non-structural proteins of Sindbis virus, nsP1, 2, 3 and 4, are produced upon cleavage of polyproteins P123 and P1234 by a proteinase residing in nsP2. We used cell free translation of SP6 transcripts to study the proteolytic activity of nsP2 and of nsP2-containing polyproteins. To generate polyprotein enzymes, a set of plasmids was made in which cleavage sites were eliminated and new initiation and termination codons introduced by *in vitro* mutagenesis. As a substrate, we used a polyprotein in which the nsP2 proteinase had been inactivated by a single amino acid substitution. All nsP2-containing polyproteins cleaved the nsP1/2 site *in trans*. However, proteinases containing nsP1 were unable to cleave the nsP2/3 site. Furthermore, only proteinases containing nsP3 could cleave the nsP3/4 site. These differences in cleavage site specificity result in a temporal regulation of processing *in vivo*. At 1.7 h post infection P123 and nsP4 accumulated and only small amounts of P34 were found. However, at 4 h post infection P123 was processed rapidly and P34 was produced rather than nsP4. Since nsP4 is thought to be the viral RNA polymerase, the temporal regulation of the nsP4/P34 ratio may be responsible for the temporal regulation of RNA synthesis.

Key words: alphavirus/non-structural proteins/polyprotein/proteinase

Introduction

Sindbis virus (SIN) is an enveloped, plus-stranded RNA virus belonging to the genus alphavirus of the family Togaviridae. The 11.7 kb genome, which is capped and polyadenylated, acts as a messenger for the production of four non-structural proteins (nsP1, nsP2, nsP3 and nsP4, ordered from the NH_2-terminus), which are thought to form the viral transcriptase–replicase complex. During replication, the genomic RNA is transcribed into full-length minus-strand RNA, which in turn serves as a template for the synthesis of new genomic RNA and 26S subgenomic RNA. The latter RNA species is the messenger for the structural proteins (Strauss and Strauss, 1986). The synthesis of minus-strand RNA ceases 3–4 h post infection, while the production of genomic and subgenomic RNA continues throughout the infectious cycle (Sawicki and Sawicki, 1980; Sawicki *et al.*, 1981a,b). The mechanism by which minus-strand synthesis is switched off is unknown.

The non-structural proteins are made as two polyprotein precursors (Strauss *et al.*, 1983, 1984). For both precursors translation starts at the same initiation codon. In most instances translation stops at an opal termination codon downstream of the nsP3 gene, resulting in polyprotein P123 (200 kd). However, at a rather high frequency, read-through of the termination codon occurs leading to the synthesis of a larger polyprotein (250 kd) which includes nsP4 (P1234).

Although the functions of the non-structural proteins have not been fully elucidated, the characterization of temperature sensitive mutants and protein sequence comparisons have provided important clues. nsP4 is hypothesized to be the actual RNA polymerase (Kamer and Argos, 1984; Hahn *et al.*, 1989a). nsP1 is involved in minus-strand synthesis (Hahn *et al.*, 1989b) and also contains a methyl-transferase activity needed for the capping of viral RNAs (Mi *et al.*, 1989). nsP2 is required for the synthesis of the 26S subgenomic RNA and for the shut-off of minus-strand synthesis (Hahn *et al.*, 1989b). Interestingly, nsP2 also contains the proteinase responsible for the processing of the non-structural polyproteins (Ding and Schlesinger, 1989; Hahn *et al.*, 1989b; Hardy and Strauss, 1989), and thus in principle could regulate the synthesis of minus-strand and subgenomic RNA through proteolytic cleavages. The proteolytic activity was localized to the carboxy-terminal half of nsP2. This domain shows a limited sequence similarity to the papain-like thiol proteinases, implicating Cys 481 as the active site residue (Hardy and Strauss, 1989).

Processing of P123 *in vitro* was shown to be sensitive to dilution, indicating that cleavage of this protein occurs predominantly in a bimolecular reaction (*in trans*); in contrast, processing of P12 was not influenced by dilution, strongly suggesting autoproteolysis (cleavage *in cis*) (Hardy and Strauss, 1989). *In vivo*, the kinetics of processing indicated that at 3–4 h after infection, P123 is first cleaved to P12 and nsP3, followed by cleavage of P12 (Hardy and Strauss, 1988). Paradoxically, however, the elimination of the nsP2/3 cleavage site by site specific mutagenesis did not influence the cleavage at the nsP1/2 and nsP3/4 sites, as studied *in vitro*, but the elimination of the nsP1/2 site prevented the cleavage at the nsP2/3 site and resulted in the accumulation of P123, i.e. cleavage of the nsP1/2 site appears to be essential for initiation of the processing pathway (Shirako and Strauss, 1990). These conflicting observations can be rationalized by postulating that (i) there are proteolytically active polyproteins containing the nsP2 region; (ii) the different polyprotein proteinases differ in their preferences for the three cleavage sites; and (iii) at 3–4 h after infection proteinases with a preference for the nsP2/3 site predominate. In this paper we show by cell free translation of synthetic transcripts that the different polyproteins are indeed active proteinases that differ in their cleavage site

preferences. Furthermore, we provide evidence that these differences result *in vivo* in a temporal regulation of polyprotein processing. The possible consequences for regulation of viral replication are discussed.

Results

Construction of cDNA clones for expression of nsP2-containing polyproteins

In vitro transcription of cDNA clones, followed by *in vitro* translation, provides a powerful approach to the study of the processing of viral polyproteins (Ympa-Wong and Semler, 1987). We have employed this strategy to determine the *trans*-cleavage activities of SIN nsP2 and of polyproteins containing nsP2. For this purpose we constructed a set of cDNA clones that when transcribed and translated *in vitro* would give rise to nsP2-containing polyproteins that could be used either as enzymes or substrate in *trans*-cleavage assays. Construction of these clones required creation of new initiation or termination codons, mutagenesis of cleavage sites to render them non-cleavable, mutagenesis of the enzymatic domain to inactivate the proteinase, and replacement of the opal codon following nsP3 with a serine codon. This set of clones and the terminology used is illustrated schematically in Figure 1A and described in more detail below.

Since nsP2 is formed by proteolytic cleavage of a polyprotein, the nsP2 gene does not have an initiation or termination codon. In plasmid pToto.2, we have provided the nsP2 gene with an initiation codon preceding the Ala1 codon, and an amber stop codon immediately downstream of the 3′ terminal Ala807 codon. The gene is located downstream of an SP6 RNA polymerase promoter and the 5′-terminal noncoding 60 nucleotides of the SIN genome, such that transcription with SP6 polymerase leads to an RNA with the authentic SIN leader immediately coupled to the nsP2 gene. *In vitro* translation of such transcripts resulted in the synthesis of an 80 kd polypeptide, which in SDS–PAGE gels comigrated with the authentic SIN nsP2 protein (Figure 1B, lane 3). To distinguish this protein product from the nsP2 derived by normal proteolytic processing of polyproteins, we will refer to it as N2.

To produce non-cleavable polyproteins containing nsP2, we took advantage of observations that the 1/2, 2/3 and 3/4 cleavage sites can be eliminated by changing them from Gly-Ala-Ala, Gly-Ala-Ala and Gly-Gly-Tyr, respectively, to Glu-Ala-Ala, Glu-Ala-Ala and Gly-Val-Tyr, respectively (Shirako and Strauss, 1990; R.J.de Groot, unpublished results). The series of cDNA plasmids illustrated in Figure 1A when transcribed and translated *in vitro* yield all possible nsP2-containing polyproteins. *In vitro* translations of SP6 transcripts derived from these plasmids are shown in Figure 1B. To distinguish these non-cleavable polyproteins from those produced by translation of wild-type RNA, we will refer to them with an N rather than a P. Thus, for example, N123 refers to the polyprotein containing the sequences of non-structural proteins 1, 2 and 3 in which the cleavage sites have been eliminated by mutagenesis, whereas P123 is the polyprotein translated from wild-type RNA.

To study the *trans*-cleavage activities of these nsP2-containing polyproteins, a substrate was required. In previous experiments, we have used cDNA clones containing deletions in the protease domain of nsP2 gene to produce

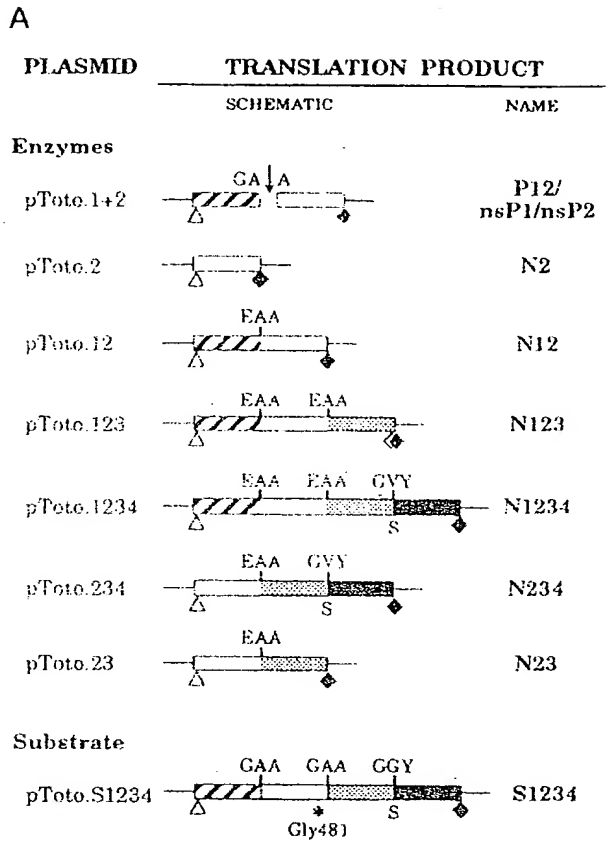


Fig. 1. (A) Schematic representation of the constructs used for expression studies. The nomenclature used for the plasmids and for the expression products obtained are indicated to the left and right, respectively. In the diagrams the non-structural protein coding regions are depicted as boxes with different shading used for each protein; the cleavage sites, either unaltered or mutagenized, are shown in the one letter amino acid code. Initiation codons are indicated by open triangles. The white diamond indicates the opal codon read through with high efficiency, which is replaced by a serine codon in the constructs pToto.1234, pToto.234 and pToto.S1234 as indicated; the solid diamonds depict termination codons resulting in efficient termination. The asterisk indicates the Cys481 to Gly substitution. (B) *In vitro* translation of SP6 transcripts derived from constructs shown in A. Rabbit reticulocyte lysate was supplemented with [^{35}S]methionine and RNA and incubated for 1 h at 30°. Translates were analyzed by 7.5% SDS–PAGE. A translate of SIN strain HR RNA served as a marker (lane 1). The positions of the SIN non-structural proteins and their precursors are indicated.

truncated polyproteins that are unable to process by themselves (Hardy and Strauss, 1989). However, these deletions may affect the overall conformation of the polyprotein and thereby influence the results of a *trans*-cleavage assay. Recently, we have found that by changing Cys 481 into Gly the proteolytic activity of nsP2 is completely abolished (R. Levinson, R.J. de Groot and J.H. Strauss, unpublished results). A plasmid was constructed in which the Cys to Gly substitution in nsP2 was combined with the opal to Ser substitution at the C-terminus of nsP3 (position 1897 of the SIN non-structural open reading frame; Li and Rice, 1989) (Figure 1). Transcription of this plasmid, pToto.S1234, followed by translation *in vitro*, resulted in the synthesis of the uncleaved 250 kd precursor (Figures 2 and 3). This product, designated S1234, was considered an ideal substrate, since the single amino acid Cys to Gly substitution is less likely to change the folding of the polyprotein than deletions in the nsP2 gene.

Trans-cleavage assays

For *trans*-cleavage assays, the nsP2-containing polyproteins used as enzymes were synthesized by *in vitro* translation using unlabeled methionine, while the substrate S1234 was radioactively labeled by performing the translation in the presence of [^{35}S]methionine. The translation reactions were stopped by adding cycloheximide and excess unlabeled methionine, after which the enzymes and substrate were incubated together. Translates of wild-type SIN genomic RNA were used as positive controls. Incubation of S1234 with the wild-type translate resulted in the production of all four non-structural proteins as well as of the polyproteins P12, P123 and P34 (Figure 2, lane 2). P12 and nsP3 were the most abundant products, indicating that cleavage occurred predominantly at the 2/3 and 3/4 sites. After incubation of S1234 with N123 or N1234 as enzymes, only the products P123, P23, nsP4 and nsP1 were found. Apparently, N123 and N1234 can cleave only the 1/2 and 3/4 sites and are unable to cleave the 2/3 site. Visual inspection of the autoradiogram suggested that most S1234 was converted into P123, indicating that the 3/4 site was predominantly cleaved (Figure 2). To assess the difference in the efficiency of cleavage of the 1/2 and 3/4 sites more precisely, we determined the amounts of input S1234 (lane labeled 'Blank') and the cleavage products P123 and P23 (lane '+N123') by densitometry (P234 was not detected). An endogenous reticulocyte protein labeled during translation served as an internal control for the amount applied to each lane. N123 cleaved 90–95% of the substrate at the 3/4 site, whereas only 35–38% of the 1/2 sites were cleaved. nsP4 was underrepresented in Figure 2. The amounts of nsP4 found in translates varied between experiments for unknown reasons. Loss of nsP4 was not always observed and appeared to depend on storage conditions and the batch of reticulocyte lysate used (not shown).

Upon incubation of S1234 with the enzymes N23 and N234, all four non-structural proteins were found, in addition to the polyproteins P123, P12, P23, P34 and P234 (Figure 2). Thus these enzymes are able to cleave all three sites. In contrast, processing of S1234 by N12 yielded only low amounts of nsP1 and P234 (Figure 3). Apparently, cleavage at the 2/3 and 3/4 sites did not occur, and enzyme N12 is able to cleave only the 1/2 site.

A translation mixture containing nsP1, nsP2 and P12

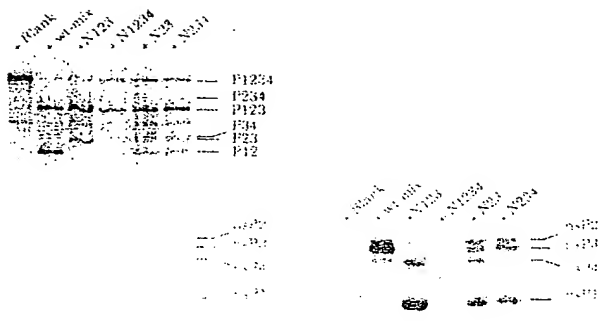


Fig. 2. *Trans*-cleavage assay: processing of the S1234 substrate by N123, N1234, N23 and N234 *in vitro*. ^{35}S-labeled S1234 was prepared as described in Figure 1. The polyproteins used as enzymes were synthesized similarly except that unlabeled methionine was used. Translation reactions were stopped by adding cycloheximide and excess unlabeled methionine after which the enzymes and substrate were mixed at a 1:1 (v/v) ratio and allowed to incubate for an additional 2 h at 30°. The cleavage products were analyzed by 7.5% SDS–PAGE. Translates without added RNA (Blank) and with SIN strain HR viral RNA (wt-mix) served as negative and positive controls, respectively. The right hand panel shows a longer exposure of the bottom half of the gel.

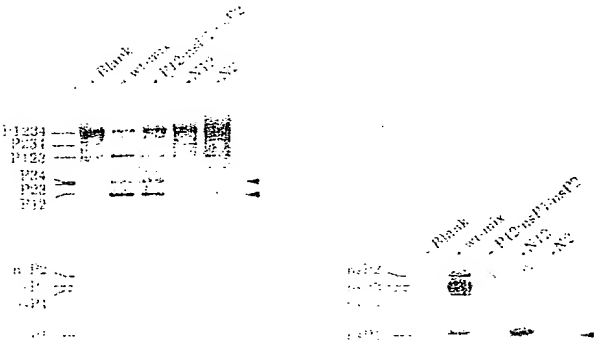


Fig. 3. *Trans*-cleavage assay: *in vitro* processing of the S1234 substrate by N2, N12 and a translate containing P12, nsP1 and nsP2 (derived from pToto.1+2). The experimental procedures were as in Figure 2. The right hand panel shows a longer exposure of the bottom half of the gel. The arrows point to P34, P12 and nsP1 in the N2 digest of S1234.

derived from pToto.1+2 (Figure 1B, lane 2) cleaved S1234 predominantly at the nsP2/3 cleavage site: the major products observed were P12 and P34. The 1/2 site was cut less efficiently, and cleavage of the 3/4 site was not detected (Figure 3). If we assume that P12 has the same cleavage specificity as N12, these findings indicate that nsP2 produced by cleavage of P12 is able to cleave the 2/3 site but not the 3/4 site, with possible activity at the 1/2 site unresolved. Essentially the same results were obtained upon incubation of S1234 with N2 derived from pToto.2, as this enzyme was able to cleave the 1/2 and 2/3 sites but not the 3/4 site. However, the proteolytic activity of N2 was surprisingly low (Figure 3). To rule out the possibility that these results were caused by mutations introduced into pToto.2 and pToto.12 during cloning, the nsP2 genes in constructs pToto.12 and pToto.2 were sequenced, and the sequence in both was identical to that in the parental cDNA clone pToto1101. In addition, the 2029 bp *Eco*RI–*Pst*I fragments (nucleotides 1920–3949 of the SIN genome; 83% of the nsP2 gene) of

pToto.2 and pToto.12 were used to replace the corresponding fragments in pToto1101. The non-structural polyproteins translated from transcripts of the resulting plasmids were processed normally (data not shown), and thus the protease domain appears to be normal. The reason for the poor proteolytic activity of N2 remains unclear. The most likely explanation is that nsP2, normally produced as part of a polyprotein, requires the flanking protein sequences to adopt its correct (i.e. proteolytically active) conformation. Alternatively, nsP1 could act as a cofactor for optimal cleavage by nsP2.

The results of the *trans*-cleavage assays are summarized in Table I. The specificities of the various polyproteins can be described with the following rules: (i) the presence of nsP4 sequences does not affect the specificity of the proteinase, i.e. polyproteins with or without nsP4 had the same cleavage specificity; (ii) if nsP1 is present in the polyprotein, the proteinase is unable to cleave the 2/3 site; proteinases lacking the nsP1 moiety cleave the 2/3 site efficiently; (iii) the presence of nsP3 in the proteinase is required for cleavage of the 3/4 site.

Temporal regulation of the [nsP4]/[P34] ratio

The differences in site-preference of the proteinase precursors, observed *in vitro*, prompted us to consider the possibility of post-translational regulation of processing *in vivo*. In particular, our data predicted a temporal regulation of the nsP4 to P34 ratio. The rationale for this is that very early in infection, P123 and P1234 are expected to be abundant. Since these enzymes predominantly cleave the 3/4 site and do not cleave the 2/3 site, nsP4 would be generated rather than P34. However, later in infection, enzymes with a preference for the 2/3 site (Table I) will have accumulated (Hardy and Strauss, 1988). Rapid cleavage of the 2/3 site would not only prevent accumulation of P123 and P1234, but would also eliminate the proteinases capable of cleaving the 3/4 site (P123, P1234, P23 and P234), thus resulting in the accumulation of P34 rather than nsP4.

To test this hypothesis, primary chicken embryo cells infected with SIN strain HR were labeled during a 5 min pulse with [^{35}S]methionine early (1 h 45 min) and late (4 h) after infection, followed by a chase with excess unlabeled methionine (it should be noted that under these conditions ~15 min are required to synthesize a complete SIN polyprotein, and thus the first part of the chase period is involved with completion of initiated labeled chains; Hardy and Strauss, 1988). Cell lysates were immunoprecipitated with a rabbit antiserum specific for nsP4 (Hardy and Strauss, 1988). Early in infection, P1234 was predominantly processed to produce nsP4, as predicted (Figure 4, early). P34 accumulated to some extent during the 5 min chase, but then declined, most likely to yield nsP3 and nsP4. The amount of P1234 present could not be determined because of the presence of a 250 kd host protein which precipitated non-specifically. Conversely, at 4 h after infection P34 accumulated rather than nsP4 (Figure 4, late). The 250 kd host protein was not present, probably due to effective shut-off of host protein synthesis, allowing visualization of P1234. Since P1234 was not present in large amounts, cleavage at the 2/3 site occurred either cotranslationally or immediately after termination of translation (Figure 4, late). During the chase P34 was converted into a form with slightly lower mobility, most likely due to phosphorylation (Hardy and Strauss, 1988; Peränen *et al.*, 1988; G.Li, M.W.LaStarza, W.R.Hardy, J.H.Strauss and C.M.Rice, in preparation).

Table I. Enzymes that cleave each of the three different sites in S1234

Site 1/2	Site 2/3	Site 3/4
nsP2	nsP2	P123
P12	P234	P1234
P123	P23	P234
P1234		P23
P234		
P23		

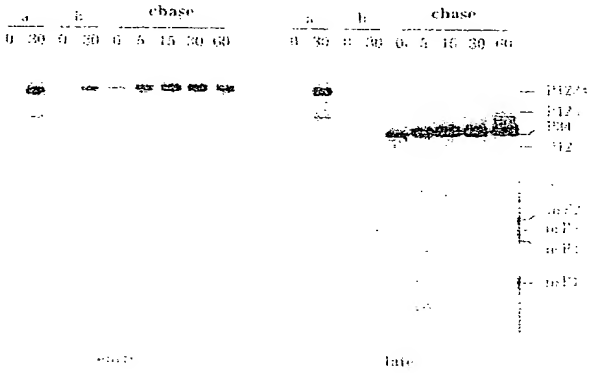


Fig. 4. Synthesis of P34 and nsP4 early and late in SIN infection. Confluent monolayers of secondary chicken embryo cells were infected with SIN strain HR at an m.o.i. of 50 p.f.u./cell. After 1 h incubation at 37°C, the inoculum was removed and replaced by Eagle's medium, containing 5 μM methionine. At 1 h 45 min p.i. ('early') or 4 h p.i. ('late') the viral proteins were labeled with [^{35}S]methionine during a 5 min pulse and then either lysed immediately or after a chase with excess unlabeled methionine for the periods of time indicated (in min). The lysates were immunoprecipitated with a non-specific rabbit antiserum raised against nsP4. Immunoprecipitates were analyzed by 7.5% SDS–PAGE. The following controls were included: (a) mock-infected cells were lysed after a 5 min pulse labeling and a 0 min or 30 min chase, and subjected to immunoprecipitation with αnsP4 serum; (b) SIN infected cells were lysed after a 5 min pulse labeling and a 0 or 30 min chase and subjected to immunoprecipitation with αnsP4 pre-immune serum. An *in vitro* translate of SIN strain HR viral RNA served as a marker. The positions of the SIN non-structural proteins and their precursors are given.

To study the processing of nsP2-containing polyproteins at the two different time-points, immunoprecipitations were also performed with a rabbit antiserum directed against nsP2 (Figure 5). Early in infection, P123 accumulated during the first 15 min of chase but declined thereafter, whereas the amount of nsP2 increased steadily during the chase (Figure 5, early). In contrast, P123 did not accumulate late in infection, but appeared to be processed rapidly to generate P12 and nsP2 (Figure 5, late).

Discussion

Many RNA viruses express their genetic information by synthesis of polyproteins that are post-translationally cleaved to generate the functional viral gene products. Host proteinases may be involved in the case of structural proteins that mature in subcellular organelles, but in the case of proteins processed in the cytosol the proteinases responsible are encoded by the virus itself (Rice and Strauss, 1981; Wellink and van Kammen, 1988). Alphaviruses produce both

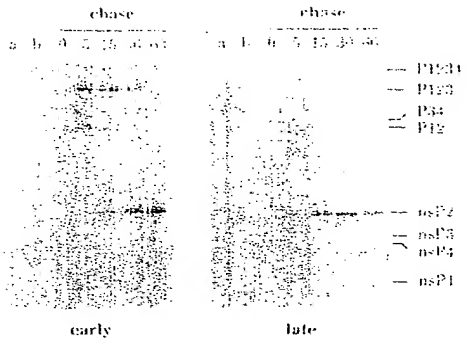


Fig. 5. Synthesis of nsP2 and nsP2-containing precursors early and late in SIN infection. The samples described in Figure 4 were immunoprecipitated with a monospecific rabbit antiserum raised against SIN nsP2. Material from an equivalent number of cells was loaded in each lane, and the samples were analyzed in the same gel. A 10 day exposure of the autoradiogram is shown for the panel labeled 'early' and a 1.5 day exposure for the panel labeled 'late'. The following controls are included: (a) mock-infected cells, lysed after a 5 min pulse labeling and a 30 min chase, subjected to immunoprecipitation with αnsP2 serum; (b) SIN infected cells, lysed after a 5 min pulse labeling and a 30 min chase, subjected to immunoprecipitation with αnsP2 preimmune serum.

their structural and non-structural proteins by processing of polyprotein precursors. Whereas the proteolytic cleavage of the structural polyprotein has been studied in considerable detail (Hahn *et al.*, 1985; Melancon and Garoff, 1986, 1987; Schlesinger and Schlesinger, 1986; Strauss and Strauss, 1986), only recently have details about the processing of the non-structural proteins become known. These studies on the non-structural proteins have mostly used SIN, and have been aided by the availability of a full-length cDNA clone (Rice *et al.*, 1987) and monospecific antisera raised against each of the non-structural proteins (Hardy and Strauss, 1988). It has now been firmly established that all three cleavage sites in P1234 are cut by a proteinase residing in the C-terminal half of nsP2 (Ding and Schlesinger, 1989; Hardy and Strauss, 1989; Shirako and Strauss, 1990).

Here, we have studied the proteolytic activities of nsP2-containing polyproteins by using *in vitro* mutagenesis techniques and cell free translation of synthetic transcripts. Our data show that all nsP2-containing polyproteins are proteolytically active and can cleave the 1/2 site *in trans*. However, the precursors differ strikingly in their cleavage site preference with respect to the 2/3 and 3/4 sites. Polyproteins containing nsP1 were unable to cleave the 2/3 site *in vitro*, whereas the proteinases lacking nsP1 cleaved the 2/3 site very efficiently. Furthermore, the presence of nsP3 in the proteinase appeared to be required for cleavage of the 3/4 site. An analogous situation exists for poliovirus in that the P2 polyprotein is cleaved *in trans* by the 3C proteinase (Ypma-Wong and Semler, 1987) but the P1 polyprotein can only be cleaved by the 3CD proteinase precursor (Jore *et al.*, 1988; Ypma-Wong *et al.*, 1988). The mechanism by which the SIN proteinases change their cleavage site preference is unknown. It is possible that after nsP1 is removed, the proteinase refolds and assumes an altered conformation that then enables it to cleave the 2/3 site. Alternatively, the nsP1 sequences may simply prevent cleavage of the 2/3 site by steric hindrance. The sequence of the cleavage site could also be important: the 3/4 cleavage site is Gly-Gly-Tyr, while the 1/2 and 2/3 sites are both Gly-Ala-Ala.

Processing of P123 *in vitro* is dilution sensitive, indicating that at least the initial cleavage between nsP1 and nsP2 occurs in a bimolecular reaction (Hardy and Strauss, 1989; Shirako and Strauss, 1990). Apparently, it is this dependence on *trans*-cleavage combined with the fact that the proteinases differ in their cleavage site preference that allows the virus to regulate post-translationally the synthesis of the non-structural proteins and their precursors. Our present view on the early events in SIN infection is as follows (Figure 6). After its release in the cytoplasm, the genomic RNA is translated into P123 and P1234. These proteinases preferentially cut the 3/4 site, resulting in the accumulation of P123 and nsP4 (Figure 6A). The 1/2 site is also cut but with lower efficiency than the 3/4 site (Figure 6B). Cleavage of the 1/2 site, however, unleashes proteinases with a strong preference for the 2/3 site. As these latter proteinases accumulate in the cytoplasm, the cleavage of the polyproteins is gradually redirected from the 3/4 site to the 2/3 site, generating P34 and large amounts of nsP3 and P12. The P12 precursor cleaves *in cis* to generate free nsP1 and nsP2 (Hardy and Strauss, 1989) and thereby adds to the proteolytic activity directed against the 2/3 site. Finally, at 3−4 h after infection a situation is reached in which nsP2 is present in such high concentrations that cleavage of the 2/3 site occurs either cotranslationally or immediately after termination of translation, thereby simultaneously preventing accumulation of P123, eliminating the enzymes that can cleave the 3/4 site, and generating P34 rather than nsP4 (Figure 6C). The fact that at 3−4 h after infection P34 is abundant and nsP4 is present in only low amounts (if at all), had previously led to the hypothesis that P34 is the active protein species rather than nsP4 (Hardy and Strauss, 1988; Strauss *et al.*, 1988). Our present results suggest that both protein species are functional, but at different times in infection.

It seems highly likely that the temporal regulation of the non-structural proteins is important for the development of the virus life cycle. For one, polyproteins like P123 could in principle perform functions required early in infection that cannot be performed by the final endproducts. But even more intriguing is the temporal regulation of the ratio of nsP4 to P34. Several observations indicate that nsP4 is the actual RNA polymerase. It contains the Gly-Asp-Asp motif found in the polymerases of various plant and animal viruses (Kamer and Argos, 1984). Moreover, certain *ts* mutations in nsP4 result in a total termination of RNA elongation at the non-permissive temperature (Barton *et al.*, 1988; Hahn *et al.*, 1989a). It is therefore tempting to speculate that the regulation of the nsP4 to P34 ratio is correlated with a temporal regulation of RNA synthesis, e.g. regulation of minus-strand synthesis, which normally ceases at 3−4 h after infection (Sawicki and Sawicki, 1980; Sawicki *et al.*, 1981a, b), of subgenomic RNA synthesis, or both. Since nsP4 is predominantly made early in infection, this protein could be a polymerase required for minus-strand synthesis, while P34 could be required for the synthesis of the subgenomic RNA species. In this case, competition between nsP4 and P34 for association with replicase complexes could provide a plausible explanation for the shut-off of minus-strand synthesis at 3−4 h after infection. This hypothesis is supported by results obtained for mutant *ts*17. At the restrictive temperature, this mutant shows aberrant processing of the 2/3 site, decreased synthesis of subgenomic RNA, and continuous minus-strand synthesis (Hardy *et al.*,

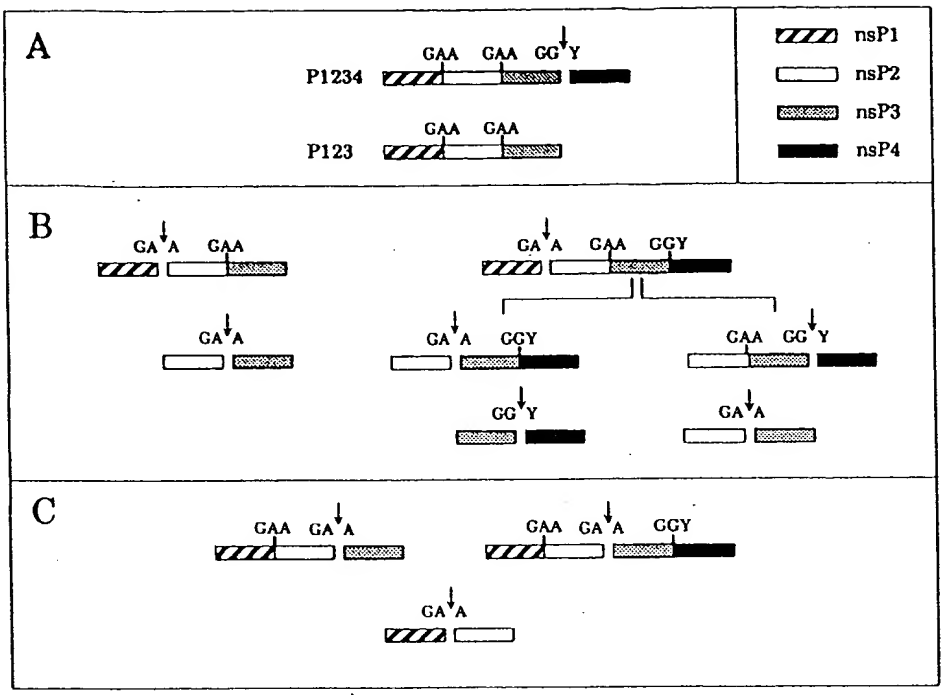


Fig. 6. Cleavage pathways of the SIN non-structural polyproteins: a model for the temporal regulation of cleavage. (A) Very early in infection cleavage of the 3/4 site is presumed to be the major cleavage pathway resulting in accumulation of P123 and nsP4. (B) P123 and P1234 are also cut at the 1/2 site early in infection but with lower efficiency. Cleavage of the 1/2 site results in the production of proteinases (nsP2, P23 and P234) that can cleave the 2/3 site. The mixture of proteinases eventually leads to the cleavage of all three bonds and thus to the production of nsP1, nsP2, nsP3 and nsP4. (C) As infection proceeds proteinase active on the 2/3 site accumulates in the cytoplasm and pathway C becomes increasingly important until at 4 h p.i. it is predominant: the polyproteins are cut at the 2/3 site cotranslationally or immediately after translation termination, preventing accumulation of P123. Proteinases that can cleave the 3/4 site (P123, P1234, P23 and P234) are eliminated and P34 is produced rather than nsP4.

1990; Hahn *et al.*, 1989b; Sawicki and Sawicki, 1985). In accordance with our hypothesis, the *ts* lesion also results in an increased production of nsP4 at the restrictive temperature, whereas P34 is turned over or processed rapidly (Hardy *et al.*, 1990). Sawicki and Sawicki (1985) showed that at the restrictive temperature, *ts*17-infected cells can resume minus-strand synthesis after complete shut-off and this resumption occurs even when protein synthesis is inhibited by cycloheximide. On the basis of these results it was concluded that proteolytic cleavage could not explain the shut-off of minus-strand synthesis. However, it seems quite possible that after the shift to the non-permissive temperature, nsP4 could be generated from pools of P34 already present in the cell, i.e. nsP4 could be produced even without *de novo* protein synthesis and lead to the resumption of minus-strand synthesis.

The question remains open as to the function of the opal termination codon between nsP3 and nsP4, especially in view of the fact that in at least two alphaviruses this codon is replaced with a sense codon (Strauss *et al.*, 1988; Takkinen, 1986). Readthrough of the SIN opal codon occurs at a frequency of ~20% *in vitro* (R.J.de Groot, unpublished data). If one assumes that readthrough occurs with similar efficiency *in vivo*, the synthesis of P34 and nsP4 would be reduced at least 5-fold as compared with nsP1, nsP2 and nsP3. Apparently, this down-regulation is either not necessary or accomplished in a different fashion in those viruses lacking the termination codon.

In summary, we have shown that the SIN non-structural proteinases differ in their cleavage site preferences and we have provided evidence that these differences result in a temporal regulation of protein processing during infection. Several observations link the temporal regulation of the nsP4 to P34 ratio to the regulation of RNA synthesis, but direct evidence for this has not yet been obtained. We hope to address this point by studying the kinetics of viral RNA synthesis for SIN mutants in which the 3/4 cleavage site has been rendered uncleavable or more efficient by site directed mutagenesis.

Materials and methods

Plasmids, enzymes and general methods

pToto1101 is a full-length cDNA clone of the HR strain of SIN from which infectious RNA can be transcribed *in vitro* (Rice *et al.*, 1987). pToto1000.S, a derivative of the full-length clone in which the opal codon between nsP3 and nsP4 has been replaced by a serine codon (Li and Rice, 1989), was kindly provided by C.M.Rice. pToto57, containing a unique *Xba*I site at position 54 of the SIN genome, was kindly supplied by R.J.Kuhn. pGEM5-Zf was obtained frm Promega Biotech. The plasmids were introduced in *Escherichia coli* strain MC1061.1 by $CaCl_2$ transformation. *E.coli* strain BW313 was used for the preparation of uracil-containing template DNA used for *in vitro* mutagenesis (Kunkel, 1985). Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, polynucleotide kinase, the Klenow fragment of *E.coli* DNA polymerase I, and SP6 RNA polymerase were from New England Biolabs. S1 nuclease was from BRL. Modified T7 DNA polymerase (Sequenase) was from US Biochemicals. *Taq* DNA polymerase was from Promega. Both single stranded and double stranded DNA were sequenced using the T7 DNA polymerase (sequenase) dideoxy chain termination method using conditions recommended by the manufacturer (US Biochemicals). Standard DNA manipulations and cloning procedures were done according to Maniatis *et al.* (1982).

Preparation of in vitro transcripts

Plasmid DNA for *in vitro* transcription was prepared by alkaline lysis (Maniatis *et al.*, 1982). The DNA templates were linearized with *Xho*I,

purified and used for *in vitro* transcription by SP6 RNA polymerase as described previously (Hardy and Strauss, 1989).

In vitro translation

In vitro translation was carried out in a 10 μl reaction containing nuclease treated, methionine depleted rabbit reticulocyte lysate (Promega) supplemented with 10 μCi [^{35}S]methionine (>1000 Ci/mmol; Amersham), 20 μM of an unlabeled amino acid mixture lacking methionine (Promega) and 10–50 ng RNA. Incubation was for 1 h at 30°C. The reactions were stopped by adding 40 μl sample buffer (Laemmli, 1970).

For *trans*-cleavage experiments, the polyproteins to be tested for proteinase activity were prepared by *in vitro* translation as described but in the absence of [^{35}S]methionine. Instead unlabeled methionine was added to a final concentration of 20 μM. Synthesis of both substrate and enzymes was stopped by adding 6 mg/ml cycloheximide and 20 mM methionine to final concentrations of 0.6 mg/ml and 1 mM, respectively. Subsequently, enzymes and substrate were mixed at a 1:1 (v/v) ratio and allowed to incubate for an additional 2 h at 30°C. The translation products were analyzed by SDS–PAGE (Laemmli, 1970) in 7.5% polyacrylamide gels (acrylamide:*bis*-acrylamide 75:1). The gels were fixed in 50% methanol, 12% acetic acid for 30 min, impregnated with EN3HANCE (Dupont) according to the instructions of the manufacturer, and autoradiographed. Densitometry was performed using a computing densitometer (Molecular Dynamics, model 300A) and the IQuant software. The relative amount of input substrate S1234 (determined for the blank) was compared with that of the cleavage products. A [^{35}S]Met-labeled endogenous protein of the reticulocyte lysate served as an internal control to assure that similar amounts of the undigested and digested S1234 had been loaded on the gel.

Plasmid constructs

The SIN nsP2 protein is generated by proteolytic cleavage of a polyprotein precursor and therefore does not have an initiation or a termination codon. To construct a transcription/translation vector that could be used to express the nsP2 gene, the 714 bp *Pst*I–*Bam*HI fragment derived from pToto1101 (nucleotides 3949–4663 of the SIN genome), was subcloned into M13mp19. The Ala1 codon of nsP3 was converted into an amber stop codon by site directed mutagenesis (Kunkel, 1985) using the oligonucleotide primer 5'-GACGCCTAGGCTCCAACTCCATCTCT-3'. The introduced nucleotide substitutions (underlined) also resulted in the creation of an *Avr*II site. The mutated *Pst*I–*Bam*HI fragment was then excised from RF DNA and inserted into an intermediate vector pSCV23, which contains the 2974 bp *Bgl*II-*Spe*I fragment of pToto1101 (nucleotides 2288–5262 of the SIN genome) (Shirako and Strauss, 1990), and which had been prepared by digestion with *Pst*I and *Bam*HI. The resulting plasmid was called p6-8-2. To generate an initiation codon at the 5' end of nsP2, the 1146 bp *Fnu*4HI fragment from pToto1101 (nucleotides 1677–2823 of the SIN genome) was treated with S1 nuclease followed by digestion with *Cla*I. This *Fnu*4HI(blunt-ended)-*Cla*I fragment (nucleotides 1680–2712), was then joined to the *Cla*I–*Avr*II fragment (nucleotides 2712–4099 of the SIN genome) from p6-8-2 and to the *Nco*I (blunt-ended with Klenow)–*Spe*I fragment from pGEM5-Zf in a three piece ligation. The resulting plasmid pGEM 8-2 contains a complete nsP2 gene, with an initiation codon (supplied by the blunt-ended *Nco*I site) preceding Ala1 and an opal stop codon at its 3' end. The reconstructed nsP2 gene was inserted downstream of the SP6 promoter and the SIN 5' non-coding leader sequence in a four piece ligation involving the 2849 bp *Xba*I (blunt-ended with Klenow)–*Sal*I (position 11087) fragment of pToto57 and the following fragments of pGEM 8-2: *Nco*I–*Ban*I (nucleotides 1675–1902), *Ban*I–*Cla*I (nucleotides 1902–2712) and *Cla*I–*Sal*I (nucleotides 2712–4104; the *Sal*I site was provided by the polylinker of pGEM 5-Zf). The nsP2 moiety of the resulting construct, pToto.2, was sequenced to ensure that no additional mutations had arisen during the manipulations. The sequence obtained was identical to that of the parental clone pToto1101.

pP1-539E/P2-806E is a derivative of pToto1101 in which the nsP1/2 and the nsP2/3 cleavage sites have been eliminated by changing them from Gly-Ala-Ala to Glu-Ala-Ala (Shirako and Strauss, 1990) (P1-539E signifies that amino acid 539 of nsP1 has been changed to Glu, etc.). pP1-539E/P2-806E was cut with *Cla*I (position 2712) and *Sal*I (position 11087) and ligated to the *Cla*I–*Sal*I fragment from pToto.2 to give plasmid pToto.12, which encodes a non-cleavable P12 polyprotein.

pToto.SA3 is a pToto1101 derivative in which the nsP3/4 cleavage site has been destroyed by changing it to Gly-Val-Tyr (R.J.De Groot, unpublished results) and in which the opal to Ser change following nsP3 has been transferred from pToto1000.S. To create plasmid pToto.1234 encoding a non-cleavable P1234 polyprotein, the *Spe*I-*Bss*HII fragment from pTotoSA3 (nucleotides 5262–9804 of the SIN genome) was cloned into pP1-539E/P2-806E. Plasmid pToto.234, encoding non-cleavable P234, was constructed by cloning the 9028 bp *Cla*I–*Xho*I fragment of pToto.1234 into pToto.2.

To construct transcription/translation plasmids for P123 and P23, an additional stop codon was created immediately downstream from the opal stop codon following nsP3 (position 5750) in order to eliminate the partial readthrough that occurs. A 458 bp fragment was produced by PCR amplification (Saiki *et al.*, 1988) using the oligonucleotides 5'-ATGACAGTAGCAAGGCTCACTTT-3' (SIN nucleotides 5207–5230) and 5'-GTCGACTATCAGTATTCAGTCCTCCTGCTCCTG-3' (nucleotides 5633–5665; nucleotide substitutions are underlined) as primers and pToto1101 as a template. The resulting fragment was cut with *Spe*I and cloned into pToto.1234 or pToto.234 which had been cut with *Spe*I (position 5262) and *Stu*I (position 10768), to produce pToto.123 or pToto.23, respectively.

In vivo labeling of non-structural proteins

Confluent monolayers of chicken embryo fibroblasts were infected at a multiplicity of 50 p.f.u./cell with Sindbis virus strain HR as described previously (Hardy and Strauss, 1988). After 60 min at 37°C [1 h post infection (p.i.)] the inoculum was removed and the monolayers were washed with phosphate buffered saline lacking divalent cations (PBS) to remove unabsorbed virus. Eagle's Minimal Essential Medium, pre-warmed to 37°C, containing 3% dialyzed fetal calf serum, 1 μg/ml actinomycin D, and 5 μM methionine (1/20 the normal concentration) was then added and incubation was continued at 37°C. At either 1 h 45 min p.i. or 4 h p.i. the medium was removed and the cells were washed with PBS, prewarmed to 37°C, to remove any residual methionine. The cells were then labeled for 5 min in Eagle's methionine free medium supplemented with 80 μCi/ml [^{35}S]methionine (>1000 Ci/mM, Amersham Corp.). After the pulse the cells were lysed, either immediately or following a chase for various times in Eagle's medium containing 2 mM methionine. The preparation of whole cell lysates and the immunoprecipitations were performed as described previously (Hardy and Strauss, 1988). The immunoprecipitated products were analyzed by electrophoresis on 7.5% SDS–polyacrylamide gels.

Acknowledgements

We thank Ellen G.Strauss for helpful comments and for assistance in the preparation of the manuscript. We thank C.M.Rice for supplying the plasmid pToto1000.S, Randy Levinson for generously providing the plasmid pToto.Cys481-Gly and Richard Kuhn for stimulating discussions and for supplying the plasmid pToto57. Furthermore, we thank Alexander van der Bliek for helpful suggestions concerning the PCR experiments and Michael Harrington for advice on densitometry. This work has been supported by Grants AI10793 and AI20612 from the National Institutes of Health and by Grant DMB8617372 from the National Science Foundation. R.J.de Groot was supported by a fellowship of the European Molecular Biology Organization (ALTF 280–1988).

References

Barton,D.J., Sawicki,S.G. and Sawicki,D.L. (1988) *J.Virol.*, **62**, 3597–3602.

Ding,M. and Schlesinger,M.J. (1989) *Virology*, **171**, 280–284.

Hahn,C.S., Strauss,E.G. and Strauss,J.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4648–4652.

Hahn,Y.S., Grakoui,A., Rice,C.M., Strauss,E.G. and Strauss,J.H. (1989a) *J.Virol.*, **63**, 1194–1202.

Hahn,Y.S., Strauss,E.G. and Strauss,J.H. (1989b) *J. Virol.*, **63**, 3142–3150.

Hardy,W.R. and Strauss,J.H. (1988) *J. Virol.*, **62**, 998–1007.

Hardy,W.R. and Strauss,J.H. (1989) *J. Virol.*, **63**, 4653–4664.

Hardy,W.R., Hahn,Y.S., de Groot,R.J., Strauss,E.G. and Strauss,J.H. (1990) *Virology*, **177**, in press.

Jore,J., de Geus,B., Jackson,R.J., Pouwels,P.H. and Enger-Valk,B.E. (1988) *J. Gen. Virol.*, **69**, 1627–1636.

Kamer,G. and Argos,P. (1984) *Nucleic Acids Res.*, **12**, 7269–7282.

Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Li,G. and Rice,C.M. (1989) *J. Virol.*, **63**, 1326–1337.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Melancon,P. and Garoff,H. (1986) *EMBO J.*, **5**, 1551–1560.

Melancon,P. and Garoff,H. (1987) *J. Virol.*, **61**, 1301–1309.
Mi,S., Durbin,R., Huang,H.V., Rice,C.M. and Stollar,V. (1989) *Virology*, **170**, 385–391.
Peränen,J., Takkinen,K., Kalkkinen,N. and Kääriäinen,L. (1988) *J. Gen. Virol.*, **69**, 2165–2178.
Rice,C.M. and Strauss,J.H. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2062–2066.
Rice,C.M., Levis,R., Strauss,J.H. and Huang,H.V. (1987) *J. Virol.*, **61**, 3809–3819.
Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Ehrlich,H.A. (1988) *Science*, **239**, 487–491.
Sawicki,D.L. and Sawicki,S.G. (1980) *J. Virol.*, **34**, 108–118.
Sawicki,D.L. and Sawicki,S.G. (1985) *Virology*, **144**, 20–34.
Sawicki,D.L., Sawicki,S.G., Keränen,S. and Kääriäinen,L. (1981a) *J. Virol.*, **39**, 348–358.
Sawicki,S.G., Sawicki,D.L., Kääriäinen,L. and Keränen,S. (1981b) *Virology*, **115**, 161–172.
Schlesinger,M. and Schlesinger,S. (1986) In W. Schlesinger (ed.), *The Togaviruses*. Academic Press, New York, pp. 317–392.
Shirako,Y. and Strauss,J.H. (1990) *Virology*, **177**, in press.
Strauss,E.G. and Strauss,J.H. (1986) In S. Schlesinger and M.J.Schlesinger (eds), *The Togaviridae and Flaviviridae*. Plenum Publishing Corp., New York, pp. 35–90.
Strauss,E.G., Rice,C.M. and Strauss,J.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5271–5275.
Strauss,E.G., Rice,C.M. and Strauss,J.H. (1984) *Virology*, **133**, 92–110.
Strauss,E.G., Levinson,R., Rice,C.M., Dalrymple,J. and Strauss,J.H. (1988) *Virology*, **164**, 265–274.
Takkinen,K. (1986) *Nucleic Acids Res.*, **14**, 5667–5682.
Wellink,J. and van Kammen,A. (1988) *Arch. Virol.*, **98**, 1–26.
Ympa-Wong,M.F. and Semler,B.L. (1987) *Nucleic Acids Res.*, **15**, 2069–2088.
Ympa-Wong,M.F., Dewalt,P.G., Johnson,V.H., Lamb,J.G. and Semler,B.L. (1988) *Virology*, **166**, 265–270.

Received on February 20, 1990; revised on April 27, 1990

JOURNAL OF VIROLOGY, Dec. 1988, p. 4586–4593
0022-538X/88/124586-08$02.00/0

Vol. 62, No. 12

Poliovirus Proteinase 3C: Large-Scale Expression, Purification, and Specific Cleavage Activity on Natural and Synthetic Substrates In Vitro

MARTIN J. H. NICKLIN,[1]† KEVIN S. HARRIS,[1] PETER V. PALLAI,[2] AND ECKARD WIMMER[1]*

Department of Microbiology, State University of New York at Stony Brook, Stony Brook, New York 11794-8621,[1] and Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, Connecticut 06877[2]

Received 30 June 1988/Accepted 29 August 1988

Proteinase 3C of poliovirus type 2 (Sabin) was expressed at 4% total protein in *Escherichia coli*. The protein was soluble and could be purified by a simple scheme. It was weakly active on the capsid precursor P1 (expressed in vitro), which contains two cleavage sites. The products of processing of P1 were 1ABC and 1D (VP1). The activity was insensitive to Triton X-100. Crude extracts of cells infected with poliovirus type 1 (Mahoney) gave strong processing and yielded 1AB (VP0), 1C (VP3), and 1D in the same assay system but were sensitive to detergent. 3C from cell extracts that was separated from its precursors resembled the recombinant proteinase in its activity. Recombinant 3C cleaved the peptide dansyl-Glu-Glu-Glu-Ala-Met-Glu-Gln-Gly-Ile-Thr-Asn-Lys-NH_2 at the Gln-Gly bond. We conclude that 3C is merely the core of the Gln-Gly-cleaving activity which processes P1 in vivo and that there is probably a hydrophobic contact between a larger 3C precursor and its P1 substrate which allows the second processing reaction: 1ABC, 1D → 1AB, 1C, 1D.

Poliovirus is a picornavirus and is the best-studied representative of the genus *Enterovirus*, which includes a variety of serious pathogens. Enteroviruses are also broadly related to the rhinoviruses, which are etiological agents of the common cold. These genera have essentially the same genomic structure, and within this group results for a single virus are mostly of general significance. The genome of poliovirus is a single ~7.5-kilobase RNA molecule that includes an open reading frame large enough to encode a 247-kilodalton polyprotein (14). The polyprotein is cotranslationally processed by at least two viral proteinases, mapping to the 3C and 2A regions of the viral genome (9, 29). Most of the processing sites are between Gln and Gly residues (14), and it has been shown (8) that the polypeptide sequence of the viral product 3C is essential for this process and furthermore that 3C releases itself from longer precursors (9) by cleavage at the two Gln-Gly bonds which flank the polypeptide. Two other polypeptides that contain all of the 3C sequence are present at high concentration in infected cell extracts. The most abundant is 3CD, which also contains the whole 3D polymerase sequence and appears to be a relatively stable precursor of both 3C and 3D. The other polypeptide, 3C′, is derived from 3CD by cleavage at a 2A-specific site (Tyr-Gly) and contains a small segment of the 3D protein. It has not been shown previously that 3C itself is catalytically active, although it releases itself from small engineered precursors (9, 12).

On the basis of a comparison of the encoded amino acid sequences of related viruses, the likely catalytic residues of 3C were suggested (1). These results were in agreement with the identification of 3C from encephalomyocarditis virus (EMCV) as a cysteine proteinase by biochemical methods (22). Site-directed mutagenesis has been used to confirm the essential function of Cys-147 and His-161 (12).

It appears that any evolutionary relationship between 3C and nonviral proteinases (notably, cysteine proteinases in particular) is remote, although it has been suggested that 3C is homologous to the trypsinlike family of serine proteases (7; J. F. Bazan and R. J. Fletterick, Proc. Natl. Acad. Sci., in press). The apparently strong specificity of 3C and its homolog toward the *P1* residue (glutamine), of itself, is a remarkable feature in a cysteine proteinase. Together, these observations indicate that the active site of 3C is a suitable target for the design of antiviral drugs that would have a minimal effect on healthy proteolysis by host enzymes.

Other workers have described expression systems for 3C, but it has been necessary to develop a *trans* assay for 3C activity to show that any purification scheme would yield active material for biochemical characterization and crystallography. Furthermore, it has been reported that 3C expressed in *Escherichia coli* is insoluble. We (18) and others (32) have reported methods for assaying Gln-Gly cleavage activity (QG-ase) in infected cell extracts by *trans*-cleavage of the capsomer precursor P1 (Fig. 3A), which we express in vitro. When infected cell extracts are used as the source of QG-ase, the products of processing of P1 are as seen in vivo, namely, P1 → 1AB, 1C, 1D. We used this assay for protein expressed in *Escherichia coli*. Ultimately, it will be more convenient to use a peptide substrate for 3C, and we demonstrate here that a peptide can be processed specifically by the recombinant enzyme.

MATERIALS AND METHODS

Bacterial culture medium. M medium was M9 medium (17) containing 2 g of NH_4Cl per liter and supplemented with 2 g of Casamino acids per liter, 10 g of glucose per liter, 20 μM ferric chloride per liter, and 0.1 g of sodium ampicillin per liter.

Buffers. The pHs of all buffers were determined at 22°C. Buffer A was 100 mM NaCl–40 mM Tris hydrochloride (pH 7.9). Buffer B was 100 mM KCl–20 mM Tris hydrochloride–10 mM $MgCl_2$–5 mM dithiothreitol (DTT)–1 mM EDTA (pH 7.9). Buffer C was 40 mM Tris hydrochloride–1 mM DTT (pH 7.9). Buffer D was 100 mM NaCl–20 mM *N*-2-hydroxy-

* Corresponding author.
† Present address: Institute of Molecular Pathology, A-1030 Vienna, Austria.

ethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES)–NaOH–1 mM EDTA–1 mM DTT (pH 7.4). Buffer E was 10 mM Tris hydrochloride–10 mM NaCl–1 mM $MgCl_2$ (pH 7.4). Buffer F was 10% (vol/vol) glycerol–100 mM KCl–10 mM HEPES–KOH–1 mM EDTA–1 mM DTT (pH 8.0). Buffer G was 140 mM KCl–20 mM HEPES–KOH–2 mM DTT–5 mM EDTA (pH 7.2). Buffer S (1×) was 12% (vol/vol) glycerol–76 mM Tris hydrochloride–4 mM Tris–25 mM DTT–1% sodium dodecyl sulfate.

Gel electrophoresis of protein samples. Samples were prepared in buffer S. All electrophoresis was performed in 0.1% sodium dodecyl sulfate. Except in the analysis of P1 processing, all gels were 10 to 20% polyacrylamide gradient gels (40 parts acrylamide to 1 part methylenebisacrylamide [bis]). Otherwise, gels were homogeneous 15% acrylamide gels (175 parts acrylamide to 1 part bis). The buffer system used has been described previously (18).

Preparation of substrate peptide dansyl-EEEAMEQGIT NK-NH_2. The substrate peptide dansyl-EEEAMEQGITNK-NH_2, of which the first 11 amino acids correspond to the cleavage site between poliovirus polypeptides 2A and 2B, was synthesized with an automatic synthesizer (SAM2; Biosearch) on methylbenzhydrylamine resin by standard peptide chemistry. The dansyl group was coupled to the peptide on the resin before HF cleavage by treating the free amino terminus with excess dansyl chloride. The peptide was purified to homogeneity by reverse-phase high-pressure liquid chromatography.

Construction of 3C expression plasmid pMN35. The T7 expression vector pAR2113 (24) and the expression strain *E. coli* BL21(DE3) (26) were kind gifts of J. Dunn and F. W. Studier (Brookhaven National Laboratory, Upton, N.Y.). Restriction endonucleases were obtained from New England BioLabs, Inc., and T4 DNA ligase was obtained from Bethesda Research Laboratories, Inc. Standard procedures were used in the construction of plasmids. Oligonucleotides were synthesized on a Microsyn 1450A apparatus (Systec Inc.). Poliovirus 3C cDNA was derived from a partial cDNA clone [pVS(2)2501] of the Sabin strain of poliovirus type 2 (28).

Preparation of bacterial paste. *E. coli* BL21(DE3) was transformed with plasmid pMN35. After 12 h of incubation at 37°C, a colony was picked to inoculate 50 ml of prewarmed Lennox broth and grown with shaking at 300 rpm at 37°C. After about 3.5 h, the culture (0.5 A_{600}) was added to 0.5 liter of M medium in a 2-liter flask and incubated as before for about 4 h, after which the A_{600} reached 0.5 again. The culture was injected into 9.5 liters of M medium and grown with stirring (900 rpm) and aeration (16 liters/min) in a 14-liter fermentor (Microferm II; New Brunswick Scientific Co., Inc.). The pH was maintained between 7.3 and 7.8 by addition of 10 M NaOH. When the A_{600} of a 1/5 dilution of culture (into 0.1 M NaCl) reached 0.7, expression was induced by adding 0.4 mM isopropyl-β-D-thiogalactopyranoside. After 2.5 h, the culture was harvested. Crushed ice (5 kg) was added to the culture with stirring. The bacteria were collected by centrifugation at 3,500 × *g* for 15 min. The pellets were suspended in 1 liter of buffer A and recentrifuged. The final pellet was stored at −80°C.

Purification of recombinant 3C. All procedures for purification of recombinant 3C were performed in a room refrigerated to 4°C, unless stated otherwise. Bacterial cell paste (15 g) was thawed and suspended to 40 ml with buffer A. Bacteria were lysed by two passages through a French pressure cell at 70 MPa. The lysate was mixed with a further 40 ml of buffer B (fraction 1) and centrifuged for 2 h at 360,000 × *g* in a Beckmann 60 Ti rotor (at 59,000 rpm) to yield a clear supernatant (70 ml), which was decanted and diluted to 200 ml with water. A 1.2-ml volume of 1 M Tris base was added dropwise with stirring (fraction 2). The final pH of a sample at room temperature was 8.3. This solution was layered on a column (180 ml) of DEAE-cellulose (Whatman DE-52) that had been preequilibrated with buffer C. The sample was eluted with buffer C. The first 130 ml of effluent was discarded, and the next 320 ml was collected. A 2.5-ml volume of 0.5 M EDTA-Na_3H and 5 ml of 0.5 M morpholinoethanesulfonic acid were added dropwise with stirring (fraction 3). Ammonium sulfate was added, with stirring, at 0.55 g/ml to fraction 3 over 15 min, and precipitation was allowed to proceed for 12 h. Precipitated protein was collected by centrifugation at 6,000 × *g* for 60 min at 8°C. The supernatant was discarded, and the pellet was carefully drained and suspended in buffer D to less than 4 ml (fraction 4B). Insoluble material was removed by centrifugation at 4,000 × *g*. To concentrate the protein, 2 g of solid ammonium sulfate was added with gentle agitation until dissolved. The mixture was allowed to stand at 0°C for 30 min, and then the precipitate was collected by centrifugation at 10,000 × *g* at 4°C for 30 min. The pellet was drained, and the precipitate was redissolved in 1 ml of buffer D (fraction 5). In a separate purification, fraction 5 was run on a column (55 by 1.6 cm) of superfine Sephadex G-75 (Pharmacia) in buffer D at 6 ml/h. Peak fractions (10.5 ml total) were identified by electrophoresis and pooled. Protein was collected by ammonium sulfate precipitation (0.5 g added per ml), followed, after 2 h, by centrifugation at 10,000 × *g* for 30 min. The final pellet (fraction 6) was redissolved to about 1 ml in buffer D. When purified, the concentration of 3C was estimated by measuring the A_{280}. The protein contains no tryptophan and seven tyrosine residues per molecule of M_r 20,000; thus, the extinction coefficient at 280 nm is about 8,400 M^{-1} cm^{-1} and the A_{280} 0.1% value is about 0.42 M^{-1} cm^{-1} (6).

Preparation and fractionation of infected HeLa cell lysate. HeLa R19 spinner cells (1.2×10^9) were infected with poliovirus type 1 (Mahoney) at 100 PFU per cell, as described previously (8). At 4 h later, the cells were swollen in 10 ml of buffer E at 0°C. Cells were broken by 15 strokes of a tight-fitting Dounce homogenizer. The homogenizer was drained and rinsed with 2 ml of the same buffer which was pooled with the lysate. The lysate (15 ml) was centrifuged at 10,000 × *g* for 30 min. The supernatant (12 ml) was collected (S-10). A 10-ml volume of S-10 was centrifuged at 300,000 × *g* for 60 min. The remaining 2 ml of S-10 was mixed with 0.5 ml of 50% (vol/vol) glycerol and stored at −80°C. The second supernatant (S-300; 9.5 ml) was mixed with 2.5 ml of 50% glycerol. A 10-ml volume of this fraction was mixed with 4 g of ammonium sulfate. The precipitate was collected by centrifugation at 10,000 × *g* for 30 min and suspended to 1.5 ml in buffer F. Insoluble material was removed by centrifugation at 10,000 × *g* for 5 min. A portion (0.2 ml) of the supernatant was loaded on a Superose 12 HR 10/30 column (Pharmacia) and eluted at 0.2 ml/min in buffer F. Fractions (0.5 ml each) were collected up to 24 ml. Each fraction between elution volumes 11 and 17 ml was concentrated to approximately 75 μl by centrifugal ultrafiltration for 2 h at 5,000 × *g* (Centricon; Amicon Corp.).

Correlated assays on P1 and peptide cleavage. Polypeptide P1 was supplied in reticulocyte lysate translation mixtures, as described previously (18), in which 40 g of pMN22 RNA per ml had been translated at 30°C for 60 min in the presence

of 1 mCi of [^{35}S]methionine per ml, followed by addition of nonradioactive methionine to 50 μM.

To test recombinant 3C with thiol-reactive inhibitors (E-64 and 1,3-dibromoacetone), the 3C (0.1 ml of a 0.5 mM solution in buffer D) was treated with fresh reducing agent (10 mM DTT), incubated at 30°C for 10 min, and then separated from free reducing agent by gel filtration into buffer G without DTT at 4°C. The effluent which did not contain low-molecular-weight thiol was pooled and diluted to 0.5 ml. To 90 μl of the enzyme solution (about 100 μM) was added 10 μl of buffer G without DTT (control) or a fresh solution of E-64 (to give 0.2 mM), dibromoacetone (to give 0.15 mM), or Triton X-100 (to give 1%). The mixtures were incubated at 30°C for 15 min. From each reaction mixture, 10 μl was withdrawn and added to 20 μl of buffer G–10 mM DTT, except for the final mixture, which was added to buffer G–10 mM DTT–1% Triton X-100. In parallel experiments, 3- and 10-μl volumes of infected S-10 were diluted to 30 μl with buffer G–10 mM DTT and a further 10 μl of infected S-10 was diluted with buffer G–10 mM DTT–1% Triton X-100. P1 translation mixture (4 μl) was then added with mixing, and the samples were incubated at 30°C for 2 h. Protein was precipitated by thorough admixture of 0.3 ml of acetone and collected by centrifugation; it was then drained and dried under vacuum. Pellets were redissolved in buffer S and analyzed by electrophoresis in 15% acrylamide gels. Gels were fluorographed with En_3Hance as described by the manufacturer (New England Nuclear Corp.) and used to expose preflashed X-ray film (A_{350} = 0.2) for 18 h at −80°C. Densitometry was performed on an LKB gel scanner.

Meanwhile, 10 μl of DTT solution (to give 10 mM) was added to the remaining 90 μl of the enzyme-inhibitor mixtures, followed by 10 μl of the peptide substrate solution (0.83 mM). These mixtures were also incubated for 2 h at 30°C. At the end of the incubation, the mixture was diluted 10-fold with water and analyzed on Mono-Q (see Fig. 5B, legend).

Quantification of protein products. The integrated absorbance of the P1, 1ABC, 1D, and, in some cases, 1C peaks was recorded. To correct for errors in loading and preparation of the reactions and deviations in lane width, the values of P1, 1ABC, and 1D were summed and then the fraction of the sum which each band represented was calculated. To convert to relative molar quantities (x_i) for each polypeptide (i), these values were divided by the number of methionine residues in each protein. In this construction, 1D contains 5, 1ABC contains 20, and P1 contains 25 Met residues.

RESULTS

Expression of proteinase 3C in BL21(DE3). The cDNA segment encoding the 3C region of poliovirus type 2 (Sabin strain) (28) was placed in the vector pAR2113 (24) under control of a T7 promoter. The initiation codon of T7 gene 10 was immediately followed by the coding sequence of 3C, followed by a termination codon. The completed plasmid, pMN35, was transferred into the expression strain, BL21 (DE3). The expression system is described in detail elsewhere (24, 26). Briefly, the plasmid contains a T7 promoter inserted into the *Bam*HI site of pBR322, which directs transcription by T7 RNA polymerase, the gene 1 product, toward the promoter of the *bla* gene. Gene 1 itself is inserted into the host chromosome, where its expression is controlled by the *lac* UV5 promoter, and is thus inducible by isopropyl-β-D-thiogalactopyranoside. Induction of gene 1 leads to rapid and overwhelming transcription of plasmid-specific RNA.

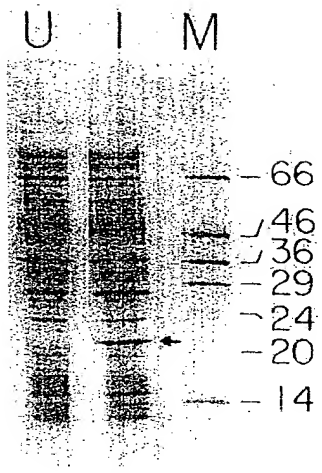


FIG. 1. Analysis on sodium dodecyl sulfate-polyacrylamide gel electrophoresis of biosynthesis of 3C in batch fermentation. Samples were withdrawn from the fermentor immediately before induction (lane U) and 2.5 h after induction (lane I) of expression with isopropyl-β-D-thiogalactopyranoside. Lanes U and I were each loaded with the protein derived from a 10-liter culture. The arrow in lane I indicates the expressed 3C. Protein was visualized by staining with Brilliant Blue R. Lane M contained molecular size markers. The numbers on the right indicate apparent molecular size (in kilodaltons).

To prepare enough cell mass to purify recombinant 3C preparatively, a benchtop fermentor was used as described in Materials and Methods. Figure 1 illustrates the induction of expression.

Purification of recombinant 3C. Since the yield of 3C was not very great, a simple purification scheme (detailed in Materials and Methods; the results are shown in Table 1 and Fig. 2) was required that could be readily repeated to yield sufficient material for X-ray crystallographic analysis. Only 5% of all the protein extracted from bacteria was not adsorbed by DEAE-cellulose (Table 1; Fig.2, compare lanes 2 and 3). Roughly 50% of this was 3C. The major contaminant was not precipitated by ammonium sulfate (Fig. 2,

TABLE 1. Purification of biosynthetic 3C from 15 g of *E. coli* BL21 (DE3) (pMN35) cell paste

Fraction no.	Description	Vol (ml)	Concn (mg/ml)	Total protein (mg)	% yield	% purity
1	Crude lysate	80	8.7[a]	690[a]	100	4
2	S-360	200	1.2[a]	240[a]	65	8
3	DEAE-cellulose effluent	320	0.03[a]	10[a]	50	50
5	Protein in second $(NH_4)_2SO_4$ precipitation	1.3	3.4[a]; 6.8[c]	4.3[a]; 8.9[c]	29[b]	90

[a] Determined by Bio-Rad protein assay (dye binding).
[b] Calculated from 3C yield in fraction 5, total protein in fraction 1, and densitometric scan of Fig. 2, lane 1.
[c] Determined by A_{280} from the estimated extinction coefficient (see Materials and Methods).

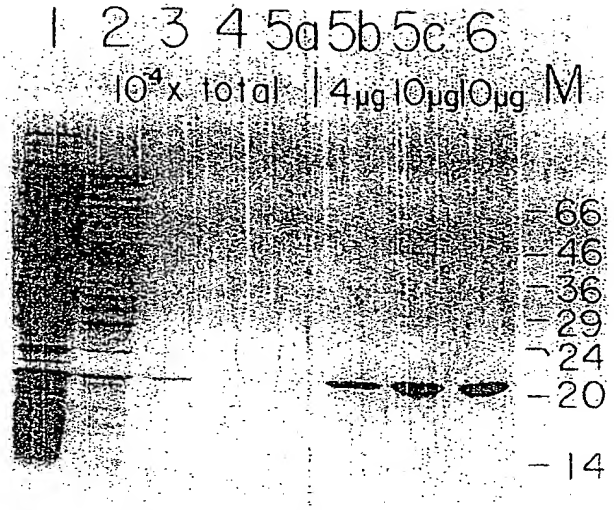


FIG. 2. Purification of recombinant 3C. For lanes 1 to 5a; a fixed proportion (1/10,000) of certain fractions (as indicated) was loaded in buffer S. Lanes 5b and 5c were loaded with 4 and 10 μg (by A_{280}), respectively, of fraction 5. Lane 6 contained 10 μg of fraction 6. Protein was visualized by staining with Brilliant Blue R. The numbers on the right (M) indicate apparent molecular sizes (in kilodaltons).

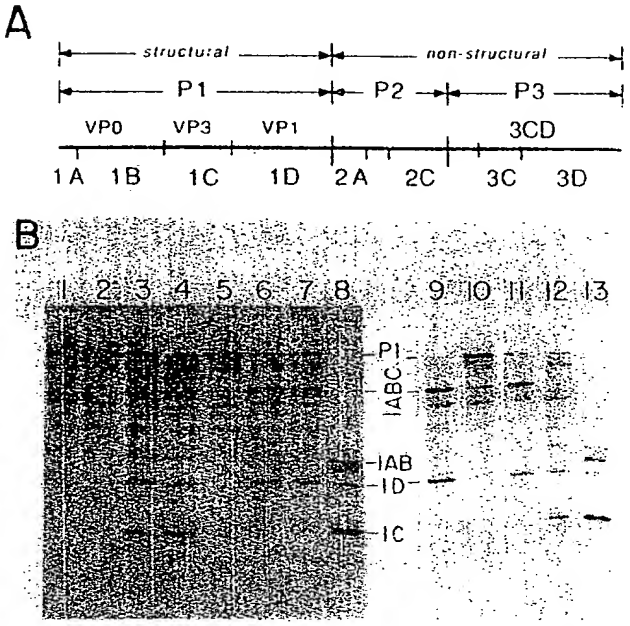


FIG. 3. Proteolytic processing of poliovirus proteins. (A) Schematic representation of the poliovirus polyprotein (heavy line). Individual polypeptides (not drawn to scale) are designated according to the nomenclature of Rueckert and Wimmer (25). The polyprotein is divided conceptually into the following three domains: P1, the capsid region which is eventually processed into the four capsid proteins 1A, 1B, 1C, and 1D (also called VP4, VP2, VP3, and VP1, respectively); P2, which yields 2A, 2B (not indicated), and 2C; and P3, which yields 3A, 3B (VPg; not indicated), 3C, 3D (the polymerase), and a variety of alternative cleavage products. Proteinase 2A (29) (and its precursors) cleaves at specific Tyr-Gly pairs (the P1-P2 site and a site in 3D). With the exception of the 1A-1B cleavage site, all other cleavages are made by 3C or its precursors (9). For a complete processing map, see Pallansch et al. (19). Note that 1AB is cleaved to 1A and 1B when the viral RNA is encapsidated and probably involves neither 2A nor 3C (2). The reaction has not been reproduced in vitro. (B, left panel) Comparison of processing of P1, expressed in vitro, by crude extracts of poliovirus-infected cells and by purified recombinant 3C. See Materials and Methods for conditions. Lanes: 1, no QG-ase; 2, no QG-ase and 1% Triton X-100; 3, 3 μl of infected S-10; 4, 10 μl of infected S-10; 5, 10 μl of S-10 and 1% Triton X-100; 6, about 25 μM 3C; 7, about 25 μM 3C and 1% Triton X-100; 8, extract of infected cells labeled for 3 h postinfection. 1AB, 1C, and 1D are indicated with arrows. (B, right panel) Effect of preincubation with thiol-reactive inhibitors on about 0.1 mM recombinant 3C for 15 min. Lanes: 9, 0.2 mM E-64; 10, 0.15 mM 1,3-dibromoacetone; 11, no inhibitor; 12, infected S-10 as in lane 4; 13, markers (as in lane 8). The proteinase was diluted to about 25 μM in the reaction mixture.

compare lanes 3 and 4). After reprecipitation from a small volume, 3C was 90% pure by densitometry (Fig. 2, lanes 5a, b, and c), and material of this quality was used in most of our experiments. Ninety-eight percent pure material was obtained by adding a single step of molecular size chromatography (Fig. 2, lane 6) with around 90% recovery. The overall recovery of protein was 30%, and one can isolate 8 mg of 3C from 15 g of wet *E. coli* cell paste.

The N-terminal sequence of the protein product was determined up to residue 10 by automated gas phase Edman degradation. The size of signal was commensurate with the sample loaded, and the data were unambiguous and did not deteriorate appreciably over the 10 cycles. The sequence determined was Met-Gly-Pro-Gly-Phe-Asp-Tyr-Ala-Val-Ala, i.e., the predicted amino terminus of 3C but with an additional methionine. Some of the purified 3C was used to raise antibodies in rabbits. The antisera strongly recognized 3C from cells infected with type 1 poliovirus (see Fig. 4A; compare lanes 31 and 32 with the recombinant 3C standards on the right of the blot).

Activity of recombinant 3C on capsid precursor P1 expressed in vitro. In preliminary experiments, crude extracts were prepared from induced BL21(DE3)(pMN35) and BL21(DE3)(pMN42), of which the latter plasmid leads to expression of the 3C protein from EMCV. No specific cleavage of poliovirus P1, expressed in vitro, was obtained with the latter expression system, but it was highly active against the EMCV 3C substrate, LVP0 (data not shown). The precursor polypeptide LVP0 of EMCV was also expressed in vitro by us, as previously described (20). Processing of poliovirus P1 by the bacterial lysate containing poliovirus 3C (the former) was qualitatively the same as that observed later with the purified protein (data not shown).

We used in vitro translation products of synthetic mRNA encoding the capsomer precursor, P1, derived from poliovirus type 1 (Mahoney) cDNA (18) to monitor QG-ase activity. After 2 h of incubation in buffer, no apparent change in the products of the translation was observed. Figure 3B, lane 1, shows the products of the control incubation. In lanes 3 and 4, infected S-10, 3 and 10 μl, respectively, was used as the source of QG-ase. As reported previously (18), processing yielded the three capsomer proteins 1AB, 1C, and 1D. It is clear that the comparative rates of processing at the two cleavage sites are similar, since all three capsomer proteins were observed even while processing of P1 was partial (lane 3). In lane 6, about 25 μM purified biosynthetic 3C was included in the incubation, giving almost complete cleavage of P1 but yielding only 1ABC and 1D, with no detectable cleavage of 1ABC to 1AB and 1C. We were able to detect a trace of 1C (identified by its mobility on gel electrophoresis)

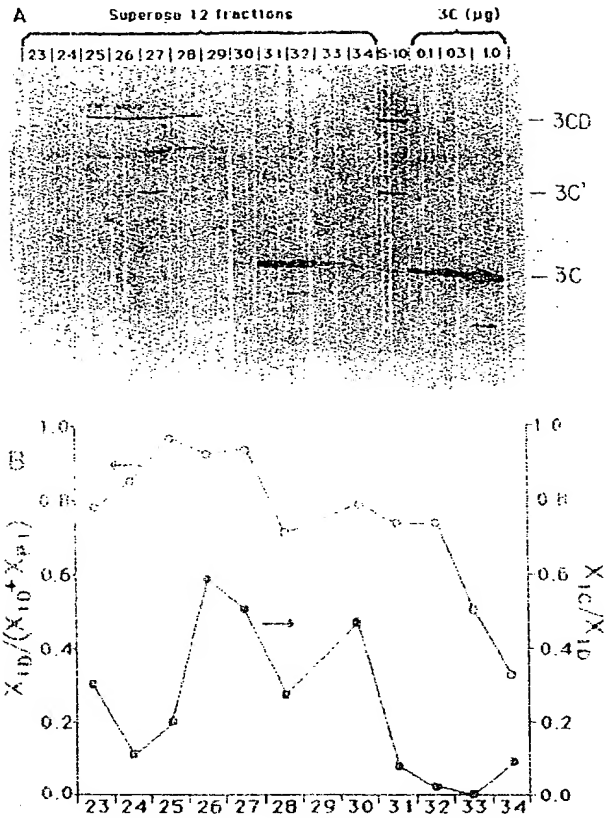


FIG. 4. Fractionation of soluble QG-ase from poliovirus-infected cells on fast-protein liquid chromatography with Superose 12 HR10/30 (bed volume, 23.6 ml). The sample size was 0.2 ml, and fractions were 0.5 ml each but were concentrated to 75 μl (see Materials and Methods). Effluent from 11 to 17 ml (fractions 23 to 34) was analyzed in detail. (A) Immunoblotting of 10 μl of each fraction (after sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was performed as described by Kräusslich et al. (16), except that the transfer was for 2 h only at 0.2 A. Rabbit anti-3C serum raised against the recombinant protein was used at a 1/300 dilution. Detection was with goat anti-rabbit immunoglobulin G coupled to alkaline phosphatase, with indolyl phosphate-Nitroblue Tetrazolium chloride as the indicator substrate system. The positions of the 3C-related proteins in the infected S-10 are labeled at the right margin. (B) A 10-μl volume of each fraction was incubated with 2 μl of P1 translation mixture for 3 h at 30°C. The gel was prepared, fluorographed, and scanned as described in the text, and the extent of cleavage (0 to 1) to yield 1D (open circles) and the molar ratio of 1C/1D (filled circles) were calculated. The arrows indicate the ordinates corresponding to open or filled circles.

at still higher 3C concentrations (data not shown). We estimate that the concentration of 3C-related proteins in the reaction mixture analyzed in lane 3 was about 1/100 of that present in the mixture analyzed in lane 6 (compare the S-10 lane with the 3C standards in Fig. 4A). Although recombinant 3C was active against P1, its activity was apparently low and restricted largely to cleavage of the 1ABC-1D bond. In separate experiments, we were able to determine the rate constant, k_{cat}/K_m, to be 80 M^{-1} S^{-1} for the interaction of 3C with P1 (data not shown).

When the same experiments were performed in the presence of 1% Triton X-100, no qualitative difference could be observed (Fig. 3B, compare lanes 6 and 7) when recombinant 3C was used as the QG-ase. Under the same conditions, the QG-ase of S-10 was completely inhibited (compare lanes 4 and 5).

Recombinant 3C and 3C from infected cells have identical activities. Recombinant 3C (unlike the QG-ase from infected cells) does not cleave P1 of poliovirus type 1 rapidly to yield 1AB, 1C, and 1D; instead, it yields only 1ABC and 1D (Fig. 3). The purified recombinant 3C is derived from the poliovirus type 2 genome and differs from the 3C in cells infected with poliovirus type 1 (Mahoney) at three amino acid positions. This should have no effect on P1 processing, however, because intertypic recombinants of poliovirus readily grow in tissue culture (23). The recombinant 3C also retains the N-terminal methionine which results from the synthetic initiation codon in the expression vector. We therefore partially purified 3C and its precursors from the soluble fraction of infected cell extracts to determine their properties. 3C has been identified as the only polioviral protein which does not sediment mainly with membranes (27). Soluble proteins in the S-300 of cell lysate were concentrated by ammonium sulfate precipitation and fractionated by molecular size. In a preliminary experiment, the size range over which the bulk of QG-ase was eluted was determined. A finer analysis was made with three assays (Fig. 4) over the range of QG-ase as follows. (i) We performed an immunoblotting analysis (Fig. 4A) with anti-3C serum. (ii) We determined the proportion of P1 that was cleaved to release 1D (Fig. 4B, open circles) in an assay. (iii) We calculated the molar ratio of 1C to 1D (x_{1C}/x_{1D}; Fig. 4B, filled circles). Since the measurements involved in calculating the third set of data were complicated, large proportional experimental errors were expected; thus, the difference in x_{1C}/x_{1D} between fractions 28 and 30 is probably not significant. The fractions which contained the peak of QG-ase activity (26 to 28) in the P1 assay did not contain 3C but did contain 3C', 3CD, and possibly other 3C-related precursors. 3CD was also detected in fraction 30 in an immunoblot treated with 10 times more concentrated anti-3C serum. It is also clear that fractions 31 and 32 contain abundant 3C and undetectable levels of 3C precursors and generate very little 1C in a P1 cleavage assay, despite extensive processing to yield 1ABC and 1D. The activity of the 3C (which is approximately 1 μM) in these fractions therefore accords well both quantitatively and qualitatively with that of recombinant 3C.

A peptide substrate specifically processed by 3C. Two procedures were used to detect the products of cleavage of the fluorescent peptide dansyl-EEEAMEQGITNK-NH_2 by recombinant 3C. In one approach (Fig. 5A and B), the peptide was subjected to extended digestion (50 μM peptide–12.5 μM 3C in buffer D for 14 h at 30°C), and all of the new UV-absorbing products which did not arise from the buffer alone were isolated by high-pressure liquid chromatography and analyzed by fast-atom bombardment mass spectroscopy (data not shown). Total conversion of the parent material (Fig. 5A, peak 1) to peaks 2 and 3 (Fig. 5B) was observed. The molecular masses of peaks 2 and 3 are consistent with the expected masses of dansyl-EEEAMEQ and GITNK-NH_2, respectively, the two products of cleavage at the Gln-Gly bond.

In separate experiments, the peptide (83 μM peptide–50 μM 3C in buffer D for 2 h at 30°C) was approximately 50% converted by recombinant 3C to a new fluorescent material with affinity for Mono-Q greater than that of the parent peptide (Fig. 5C and D and legend). After 2 h of incubation, 0.5 ml of acetone at room temperature was added to an identical reaction mixture to denature the protein and stop the reaction. The reaction mixture was dried in vacuo, and

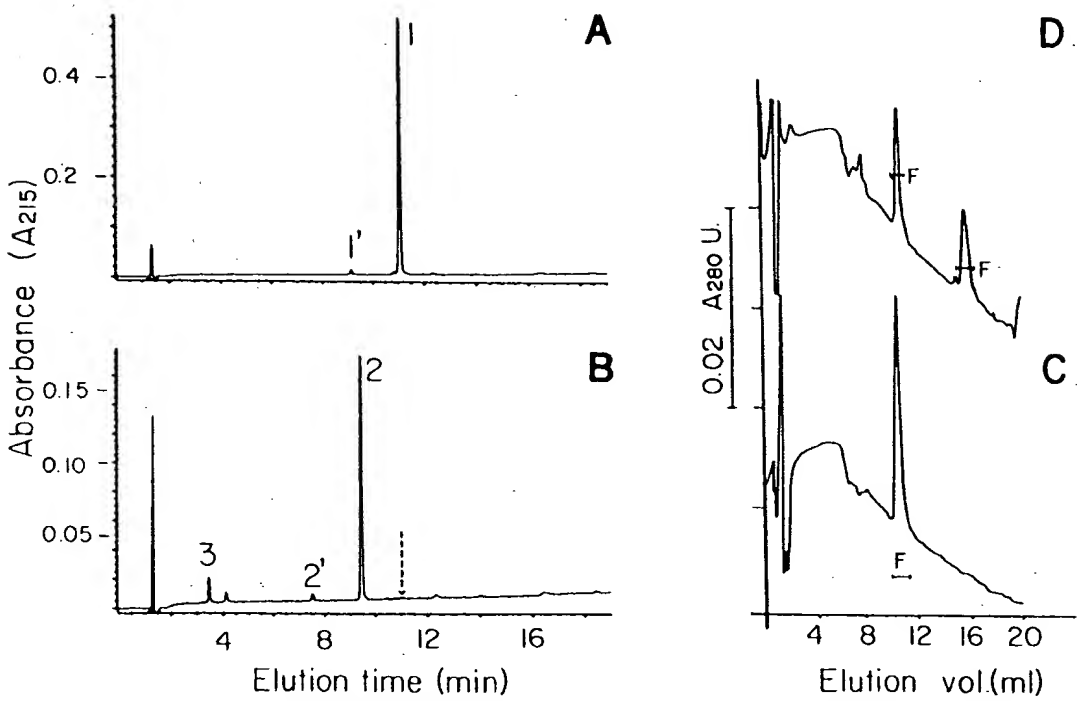


FIG. 5. Processing of the peptide dansyl-EEEAMEQGITNK-NH_2 by biosynthetic 3C. (A) Reverse-phase analysis of the substrate peptide. (B) Analysis of cleavage reaction. The peptide (50 μM) was incubated with 12.5 μM 3C for 14 h at 30°C in a final volume of 0.1 ml of buffer C. Samples were run on a Vydac C-18 column in a linear 60-ml gradient at 2 ml/min, containing 5 to 100% solution B mixed in solution A. Solution A was 0.1% trifluoroacetic acid in water, and solution B was 0.1% trifluoroacetic acid in 40% water-60% acetonitrile. The peak fractions were pooled and analyzed by fast-atom bombardment mass spectroscopy. Peak 1 yielded a molecular ion of M_r 1,610.6, as expected of the uncleaved peptide. Peak 2 yielded a molecular ion of M_r 1,098, as expected of the peptide dansyl-EEEAMEQ. Peptide 3 yielded a molecular ion of M_r 531, as expected of GITNK-NH_2. Peaks 1' and 2' yielded molecular ions of M_rs 1,626.7 and 1,114, respectively, as expected from the methionine sulfoxide derivatives of peaks 1 and 2. (C and D) Ion-exchange analysis. The peptide (83 μM) was incubated in buffer G with either no enzyme (C) or 50 μM 3C in 0.1 ml (D) for 2 h. The mixtures were diluted 10-fold and resolved on fast-protein liquid chromatography with Mono-Q HR5/5. After sample application, the column was washed with 3 ml of 20 mM bis–Tris hydrochloride (pH 6.7), followed by a 17-ml gradient from 0 to 50% 20 mM bis–Tris hydrochloride–1 M NaCl (pH 6.7) mixed with 20 mM bis–Tris hydrochloride (pH 6.7). The fluorescent fractions were detected by standing the tubes over a 330-nm transilluminator and are indicated by horizontal bars labeled F. The reaction mixture was also analyzed as described in Results.

the dry pellet was extracted with dimethylformamide and centrifuged, and the solution was transferred to a new tube for drying in vacuo. A portion of this material was analyzed by gas phase automated Edman degradation for six cycles. The sequence Gly-Ile-Thr-Asn-Lys was obtained (data not shown), with no amino acid derivative detected in cycle 6. Small contaminating peaks (running with the derivatives of Asp and Ala in cycle 1, with Gln in cycle 2, and with Ile in cycle 3) did not represent any other sequence within the substrate peptide. Together, these data show that the enzyme preparation cleaved the synthetic peptide at the Gln-Gly cleavage site.

Correlation of activity of inhibitors in P1 and peptide assays. In separate experiments, recombinant 3C (at approximately 90 μM) was pretreated with the two cysteine proteinase inhibitors L-*trans*-expoxysuccinylleucylamido(4-guanidinobutane) (E-64; 0.2 mM [3]) and 1,3-dibromoacetone (0.15 mM [11]). A control reaction contained only enzyme and buffer (see Materials and Methods). After addition of excess thiol, the mixtures were tested for activity in both the P1 and peptide assays. In the control assays, P1 was almost completely converted to 1ABC and 1D (Fig. 3B, lane 11) and the fluorescent peptide was detected under UV illumination as running almost completely in the position of the product peak, as in Fig. 5D (data not shown). The pattern of neither assay was changed (Fig. 3B, lane 9) by preincubation of 3C with E-64; hence, E-64 does not appear to inhibit the enzyme under these conditions. However, after treatment with dibromoacetone (Fig. 3B, lane 10), P1 was only weakly cleaved, whereas in fast-protein liquid chromatography roughly half of the fluorescent peptide migrated in the position of the unmodified peptide and half migrated in the position of the processed peptide (data not shown). Thus, dibromoacetone appeared to inhibit the activity of 3C in both reactions.

DISCUSSION

Expression of 3C in *E. coli*. Ivanoff et al. (12) have previously reported efficient expression (15% of total protein) of type 1 (Mahoney) poliovirus 3C from a *trp* promoter in *E. coli*, but the product was described as largely insoluble. Large amounts of internal initiation products are observed when poliovirus type 1 cDNA is used for expression of 3C (9, 12; M.J.H.N. and E.W., unpublished data). We expressed 3C derived from poliovirus type 2 (Sabin) cDNA, which does not give internal initiation in the 3C region, probably because of a single-base difference upstream of Met-27. The level of expression of the protein product of our system was only 4% by densitometry of a stained gel (Fig. 1,

compare lanes U and I), but it was largely extracted into a high-speed supernatant during purification (Fig. 2, lanes 1 and 2) and was easily purified. Other workers have also recently expressed type 2 3C, derived from the same cDNA clone, from an *E. coli* promoter (A. Nomoto, personal communication). However, like Ivanoff et al., they have found that their product is insoluble but active after renaturation.

The N-terminal methionine of our recombinant protein was unexpectedly retained (4, 30). However, as described below, the recombinant and the natural 3C seemed comparable in activity.

The processing reaction. Jackson (13) has shown in synchronized in vitro translations of EMCV RNA that the precursors of 3C are active in cleaving the capsid precursor in this system. EMCV 3C, however, is fully active upon the complex EMCV capsid precursor (20). There has previously been no clear report that 3C of poliovirus alone is a QG-ase. Here we show that recombinant 3C is proteolytically active. However, pure 3C cleaves only one of the two peptide bonds in P1 which are cleaved in vitro by the QG-ase from infected cell extracts (Fig. 3B) and are readily processed in vivo. Ypma-Wong and Semler (32) have reported previously that on translation in vitro (where the concentration of enzyme and substrate would be in the range of 10^{-8} M), the entire 3CD region must be translated for processing of P1 (synthesized from the same mRNA) to be observed, whereas the P2 and P3 regions apparently require only the 3C region itself. Our kinetic data (M.J.H.N. and E.W., unpublished data) show that 3C would need to be 2 orders of magnitude more concentrated to produce detectable processing of P1 in this system. More recent work, in which P1, P2, and P3 substrates were translated independently of the proteinase activity, has supported these findings (13a, 31a; H.-G. Kräusslich, C. Hellen, M. J. H. Nicklin, and E. Wimmer, unpublished data). It is likely, therefore, that the K_m of 3C for the P2 and P3 substrates is much lower than that for P1. We have shown that soluble precursors of 3C, when partially purified from infected cells, are highly active on P1, although 3C itself is apparently absent from these fractions. These precursors cleave P1 fully to 1AB, 1C, and 1D. The 3C from infected cells shows weak activity similar to that of recombinant 3C, yielding 1ABC and 1D. The implication of these findings is that 3C lacks a domain required for effective interaction with P1. Our observation that the QG-ase from infected cells is rendered inactive (or merely as inactive as 3C) upon the P1 substrate by inclusion of detergent in the reaction mixture, whereas high concentrations of 3C are unaffected, suggests that the interaction of the functional QG-ase and P1 is mediated by a hydrophobic interaction. In this respect it would be relatively simple to test for the involvement of the N-terminal myristoyl moiety of P1 (5, 21).

It has been reported that 3CD is not active as a polymerase (31); hence, the processing (by the QG-ase) of 3CD to 3C and 3D represents a potential regulatory step in which the P1 QG-ase is inactivated and the polymerase is activated.

Poliovirus 3C as a peptidase. Clearly, the most desirable form of assay for a proteinase is a peptidolytic assay. It would also open up a convenient route for quantitative studies of the primary specificity of the enzyme and the investigation of inhibitors. An esterolytic assay of poliovirus proteinase 2A has been reported (15), but none has been reported for 3C. The peptide sequence EEEAMEQGITN was selected to represent the QG-ase-cleavable site between polypeptides 2A and 2B, which is efficiently cleaved in vivo. The peptide synthesized was dansyl-EEEAMEQGITNK-NH_2. The reaction catalyzed by 3C yielded two peptides with molecular masses corresponding to those of the products of cleavage between Gln and Gly, and amino-terminal sequencing of all dimethylformamide-soluble peptides present in the mixture yielded a single N-terminal sequence corresponding to the C-terminal product, GITNK, as expected from the Gln-Gly specificity of the enzyme on its protein substrates. Synthetic peptides not containing a proper QG site were entirely resistant to even large amounts of 3C (P.V.P. and E.W., unpublished data). E-64, a highly specific, irreversible inhibitor of the papain superfamily (including many lysosomal cysteine proteinases) (3), apparently did not react with recombinant 3C to inhibit its activity in either the peptidolytic or the P1 cleavage assay. Dibromoacetone, on the other hand, inhibited the enzyme correspondingly. The latter inhibitor is not likely to exhibit much specificity, although it may cross-link the active-site Cys and His residues (11).

ACKNOWLEDGMENTS

We thank P. Kissel for synthesis of the oligonucleotides, M. G. Murray for uncoupling of the peptide, F. Burkhardt for high-pressure liquid chromatographic analysis, G. Hansen for mass spectroscopy, and M. Schmidt for preparation of rabbit anti-3C. We are grateful to R. Jackson, B. Enger-Valk, and B. L. Semler for communicating their data before publication. We also thank H. Toyoda, F. W. Studier, R. Cabelli, H. G. Kräusslich, and C. Hellen for stimulating discussions; A. Kameda and J. Bradley for preparing figures; L. Zywacki for help in preparing the manuscript; and C. Helmke for photography.

Gas-phase polypeptide sequencing was performed by the Center for Analysis and Synthesis of Macromolecules at Stony Brook, supported by National Institutes of Health grant RR02427 and the Center for Biotechnology, State University of New York. This work was supported in part by National Institutes of Health grants A115122 and CA28146. M.N. is fellow DRG-848 of the Damon Runyon-Walter Winchell Cancer Fund.

LITERATURE CITED

1. **Argos, P., G. Kamer, M. J. H. Nicklin, and E. Wimmer.** 1984. Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucleic Acids Res. **12**: 7251–7276.
2. **Arnold, E., M. Luo, G. Vriend, M. G. Rossmann, A. C. Palmenberg, G. Parks, M. J. H. Nicklin, and E. Wimmer.** 1987. Implications of the picornavirus capsid structure for polyprotein processing. Proc. Natl. Acad. Sci. USA **84**:21–25.
3. **Barrett, A. J., A. A. Khembhavi, M. A. Brown, H. Kirschke, C. G. Knight, C. G. Tamai, and K. Hanada.** 1982. L-*trans*-epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem. J. **201**:189–198.
4. **Ben-Basset, A., K. Bauer, S. Chang, K. Myambo, A. Boosman, and S. Chang.** 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J. Bacteriol. **169**:751–757.
5. **Chow, M., J. F. E. Newman, D. Filman, J. M. Hogle, D. J. Rowlands, and F. Brown.** 1987. Myristylation of picornavirus capsid protein UP4 and its structural significance. Nature (London) **327**:482–486.
6. **Fasman, G. D. (ed.).** 1977. CRC handbook of biochemistry and molecular biology. 3rd ed., vol. 1, 187–189. CRC Press, Inc., Boca Raton, Fla.
7. **Gorbalenya, A. E., V. M. Blinov, and A. P. Donchenko.** 1986. Poliovirus encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families. FEBS Lett. **194**:253–257.
8. **Hanecak, R., B. L. Semler, C. W. Anderson, and E. Wimmer.**

1982. Proteolytic processing of poliovirus polypeptides: antibodies to polypeptide P3-7c inhibit cleavage at glutamine-glycine pairs. Proc. Natl. Acad. Sci. USA 79:3973–3977.

9. **Hanecak, R., B. L. Semler, H. Ariga, C. W. Anderson, and E. Wimmer.** 1984. Expression of a cloned gene segment of poliovirus in *E. coli*: evidence for autocatalytic production of the viral proteinase. Cell 37:1063–1073.
10. **Hartley, B. S.** 1970. Strategy and tactics in protein chemistry. The first BDH lecture. Biochem. J. 119:805–822.
11. **Husain, S. S., and G. Lowe.** 1968. Evidence for histidine in the active site of papain. Biochem. J. 108:855–859.
12. **Ivanoff, L. A., T. Towatari, J. Ray, B. D. Korant, and S. R. Petteway, Jr.** 1986. Expression and site-specific mutagenesis of the proteinase 3C protease in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 83:5392–5396.
13. **Jackson, R.** 1986. A detailed kinetic analysis of the *in vitro* synthesis and processing of encephalomyocarditis virus products. Virology 149:114–127.

13a. **Jore, J., B. de Geus, R. J. Jackson, P. H. Powels, and B. E. Enger-Valk.** 1988. Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 *in vitro*. J. Gen Virol. 69:1627–1636.

14. **Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer.** 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature (London) 291:547–553.
15. **König, H., and B. Rosenwirth.** 1988. Purification and partial characterization of poliovirus protease 2A by means of a functional assay. J. Virol. 62:1243–1250.
16. **Kräusslich, H. G., M. J. H. Nicklin, H. Toyoda, D. Etchison, and E. Wimmer.** 1987. Poliovirus proteinase 2A induces cleavage of eucaryote initiation factor 4F polypeptide p220. J. Virol. 61: 2711–2718.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Nicklin, M. J. H., H. G. Kräusslich, H. Toyoda, J. J. Dunn, and E. Wimmer.** 1987. Poliovirus polypeptide precursors. Expression *in vitro* and processing by 3C and 2A proteinases. Proc. Natl. Acad. Sci. USA 84:4002–4006.
19. **Pallansch, M. A., O. Kew, B. L. Semler, D. R. Omilianowski, C. W. Anderson, E. Wimmer, and R. R. Rueckert.** 1984. Protein processing map of poliovirus. J. Virol. 49:873–880.
20. **Parks, G. D., G. M. Duke, and A. C. Palmenberg.** 1986. Encephalomyocarditis virus 3C protease: efficient cell-free expression from clones which link viral 5' noncoding sequences to the P3 region. J. Virol. 60:376–384.
21. **Paul, A. V., A. Schultz, S. E. Pincus, S. Oroszlan, and E. Wimmer.** 1987. Capsid protein VP4 of poliovirus is N-myristylated. Proc. Natl. Acad. Sci. USA 84:7827–7831.
22. **Pelham, H. R. B.** 1978. Translation of encephalomyocarditis virus RNA *in vitro* yields an active proteolytic processing enzyme. Eur. J. Biochem. 85:457–461.
23. **Romanova, L. I., E. A. Tolskaya, M. S. Kolesnikova, and V. I. Agol.** 1980. Biochemical evidence for intertypic genetic recombination of polioviruses. FEBS. Lett. 118:109–110.
24. **Rosenberg, A. H., B. N. Lade, D. Chui, S. Lin, J. J. Dunn, and F. W. Studier.** 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135.
25. **Rueckert, R. R., and E. Wimmer.** 1984. Systematic nomenclature of picornavirus proteins. J. Virol. 50:957–959.
26. **Studier, F. W., and B. A. Moffat.** 1986. Use of T7 RNA polymerase to direct selective, high-level expression of cloned genes. J. Mol. Biol. 189:113–130.
27. **Tershak, D. R.** 1984. Association of poliovirus proteins with the endoplasmic reticulum. J. Virol. 52:777–783.
28. **Toyoda, H., M. Kohara, Y. Kataoka, T. Suganama, T. Omata, N. Imura, and A. Nomoto.** 1984. Complete nucleotide sequences of all three poliovirus serotype genomes: implication for genetic relationship, gene function and antigenic determinants. J. Mol. Biol. 174:561–585.
29. **Toyoda, H., M. J. H. Nicklin, M. G. Murray, C. W. Anderson, J. J. Dunn, F. W. Studier, and E. Wimmer.** 1986. A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. Cell 45:761–770.
30. **Tsunawa, S., J. W. Stewart, and F. Sherman.** 1985. Amino-terminal processing of mutant forms of yeast *iso*-1-cytochrome *c*. (The specificities of methionine aminopeptidase and acetyltransferase). J. Biol. Chem. 260:5382–5391.
31. **Van Dyke, T. A., and J. B. Flanegan.** 1980. Identification of poliovirus polypeptide p63 as a soluble RNA-dependent RNA polymerase. J. Virol. 35:732–740.

31a. **Ypma-Wong, M. F., P. G. Dewalt, V. H. Johnson, J. G. Lamb, and B. L. Semler.** 1988. Protein 3CD is the major proteinase responsible for cleavage of the P1 capsid precursor. Virology 166:265–270.

32. **Ypma-Wong, M. F., and B. L. Semler.** 1987. *In vitro* molecular genetics as a tool for determining the differential cleavage specificities of the poliovirus 3C proteinase. Nucleic Acids Res. 15:2069–2088.

Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1618–1622, March 1985
Biochemistry

Murine leukemia virus protease is encoded by the *gag–pol* gene and is synthesized through suppression of an amber termination codon

(retroviral protease/amino acid sequence/translational control)

YOSHIYUKI YOSHINAKA, IYOKO KATOH, TERRY D. COPELAND, AND STEPHEN OROSZLAN

Litton Bionetics, Inc., Basic Research Program, Laboratory of Molecular Virology and Carcinogenesis, National Cancer Institute–Frederick Cancer Research Facility, Frederick, MD 21701

Communicated by David Baltimore, November 7, 1984

ABSTRACT **We have purified from Moloney murine leukemia virus (Mo-MuLV) a protease that has the capacity of accurately cleaving the polyprotein precursor $Pr65^{gag}$ into the mature viral structural proteins. Both the NH_2- and COOH-terminal amino acid sequences have been determined and aligned with the amino acid sequence deduced from the DNA sequence of Mo-MuLV by other workers. The results show that: (*i*) the protease is located at the 5′ end of the *pol* gene, and the first four amino acids are overlapped with the 3′ end of the *gag* gene; (*ii*) the fifth amino acid residue is glutamine, which is inserted by suppression of the UAG termination codon at the *gag–pol* junction; and (*iii*) the protease is composed of 125 amino acids with calculated M_r = 13,315, and the COOH terminus of the protease is adjacent to the NH_2 terminus of reverse transcriptase. The map order of the *gag–pol* gene is proposed to be 5′-p15–p12–p30–p10–protease–reverse transcriptase–endonuclease-3′.**

The internal structural proteins of murine leukemia virus (MuLV) are encoded by the group-specific antigen (*gag*) gene and synthesized as a precursor polyprotein designated $Pr65^{gag}$. In addition to the *gag* gene, all replication-competent retroviruses possess a polymerase (*pol*) and an envelope (*env*) gene which have been mapped as 5′-*gag-pol-env*-3′. Although the *gag* and *pol* genes are separated by an amber termination codon (UAG), translation of the genome-size mRNA yields, in addition to $Pr65^{gag}$, a larger precursor designated $Pr180^{gag\text{-}pol}$ (for review see ref. 1). Jamjoom *et al.* (2) suggested that the synthesis of this *gag-pol* polyprotein, which is made in amounts 4–10% of those for $Pr65^{gag}$, may be translationally controlled. Using an *in vitro* translational system and yeast suppressor tRNA, Philipson *et al.* (3) provided evidence that synthesis of $Pr180^{gag\text{-}pol}$ was enhanced by suppression of an amber termination codon.

During virus maturation $Pr65^{gag}$ is proteolytically cleaved into the final products designated p15, p12, p30, and p10. The processing is accomplished by a virion-associated protease (4), which first cleaves $Pr65^{gag}$ into $Pr27^{gag}$ (p15 + p12) and $Pr40^{gag}$ (p30 + p10), the two major intermediate cleavage products (5, 6). However, the origin of protease (viral or cellular) remained unknown. In avian retroviruses a protein designated p15 and encoded by the 3′ end of the *gag* gene has been shown to have associated protease activity (7, 8), but the *gag* proteins of MuLV were not found to cleave the precursor. Genetic studies by Traktman *et al.* (9) with conditional maturation mutants of MuLV have indicated the importance of $Pr180^{gag\text{-}pol}$ for the proteolytic processing of $Pr65^{gag}$. Levin *et al.* (10), who studied a natural *pol* frameshift mutant, confirmed and extended these observations and predicted the map position of a putative virally coded protease to be 5′ to the reverse transcriptase coding region. We have determined the NH_2-terminal sequence of the 80-kilodalton reverse transcriptase (11) derived from $Pr180^{gag}$ *pol* and located its genetic locus to begin 360 nucleotides downstream from the amber termination codon positioned at the end of the *gag* gene (12). This finding suggested to us that the *pol* gene segment upstream to the codon specifying the NH_2 terminus of reverse transcriptase may code for an approximately 14-kilodalton additional polypeptide, and we hypothesized that it may be the protease since the deduced primary and more importantly the predicted secondary structure of the putative protein resembled those of avian myeloblastosis virus and avian sarcoma virus protease (13, 14).

In this communication we report the purification and primary structure analysis of a protease from Moloney (Mo)-MuLV. These data provide the evidence that the protease is encoded by the *gag–pol* gene and is synthesized by a translational readthrough of the amber termination codon for the *gag* gene.

MATERIALS AND METHODS

Viruses. Mo-MuLV was grown in BALB/c mouse bone marrow JLS-V9 cells (MJD-54 cells) kindly supplied by K. Manly (Roswell Park Memorial Institute, Buffalo, NY). Gazdar murine sarcoma virus (Gz-MSV) was grown in HTG-2 cells (15). The viruses were purified by sucrose density gradient centrifugation and obtained from the Biological Products Laboratory, Program Resources, Inc., National Cancer Institute–Frederick Cancer Research Facility (Frederick, MD).

Assay of Protease. Gz-MSV, which itself has no protease activity and contains uncleaved $Pr65^{gag}$ as its major core protein, was the source of the polyprotein substrate. Protease activity was assayed as previously described (6).

Extraction of Protease Activity from Virus. To 50 mg of purified Mo-MuLV suspended in 2 ml of 0.13 M NaCl/0.01 M Tris·HCl, pH 7.2/0.001 M EDTA (STE buffer), 20 vol of cold acetone (−70°C) was added, and then the suspension was centrifuged at 5000 rpm for 10 min at 4°C in a Sorvall SS-24 rotor. The precipitate was dried under reduced pressure. To solubilize the protease, extraction of the acetone powder (4°C, 30 min with constant stirring) was done stepwise first with 10 ml of 0.02 M piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (Pipes), pH 7.0/5 mM dithiothreitol (PD buffer) alone, PD buffer plus 0.1 M NaCl, PD buffer plus 0.5 M NaCl, and finally PD buffer plus 2.0 M NaCl. Each aqueous extract was centrifuged at 10,000 rpm for 10 min at 4°C in a Sorvall SS-24 (SS-1) rotor. Aliquots were dialyzed against PD buffer and assayed for protease activity. Extracts

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MuLV, murine leukemia virus; Mo-MuLV, Moloney MuLV; Gz-MSV, Gazdar murine sarcoma virus; AKV, AKR mouse leukemia virus; FeLV, feline leukemia virus; BaEV, baboon endogenous virus; RP-HPLC, reversed-phase high-performance liquid chromatography.

having protease activity were pooled, concentrated by lyophilization, and saved for further purification.

NaDodSO$_4$/PAGE. Various protein materials were analyzed by discontinuous NaDodSO$_4$/PAGE (16). Specifically, to separate low molecular weight proteins, a 32-cm 8–18% polyacrylamide gradient gel was used as described (4). Visualization of proteins was by staining with either Coomassie brilliant blue R-250 or silver (17).

Ion-Exchange and Gel-Permeation Chromatography. Fractionation of protease by phosphocellulose and Sephadex G-75 chromatography was done as previously described (4).

Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC). Lyophilized samples shown to have protease activity were dissolved in saturated guanidine·HCl (Gdn·HCl), and further fractionated by RP-HPLC (18) on a μBondapak C_{18} column (Waters Associates). Five-milliliter fractions were collected and aliquots were taken for protein composition analysis on NaDodSO$_4$/PAGE and protease activity measurement.

NH$_2$-Terminal Microsequence Analysis. Semi-automated microsequence analysis was performed with a Beckman sequencer model 890C equipped with a cold trap accessory as described (19). Phenylthiohydantoin derivatives of amino acids were identified and quantitated by HPLC (20).

COOH-Terminal Sequence Analysis. Protein samples were digested with carboxypeptidase Y (21) for various time intervals and the released amino acids were quantitated on the Durrum 500 analyzer.

RESULTS

Purification of Protease. In initial studies designed to purify the Mo-MuLV protease, the proteolytic activity was first concentrated and fractionated by phosphocellulose chromatography using stepwise elution with increasing concentration of KCl. The activity that eluted at 0.3 M KCl was further fractionated by gel filtration on Sephadex G-75. Each fraction was assayed for protease activity by incubating aliquots with disrupted Gz-MSV and subsequently determining the protein pattern by NaDodSO$_4$/PAGE. Shown in Fig. 1 are the electrophoretic profiles of Sephadex G-75 fractions 20–23 after they were incubated alone or with the substrate (Gz-MuLV Pr65^{gag}). These four fractions were found to have protease activity as judged by the decrease in band intensity of Gz-MSV Pr65^{gag} and concomitant appearance of Pr40^{gag}, Pr27^{gag}, p30, and p10 bands, which were readily detectable. It is also seen by this semiquantitative assay that fractions 21 and 22 had the peak activity inasmuch as Pr40^{gag}, the proximal precursor for p30 and p10 (4), completely disappeared and was further cleaved into the final products p30 and p10. It was observed previously as well as in the present studies that, in contrast to Pr40^{gag}, the intermediate cleavage product, Pr27^{gag}, is relatively difficult to process *in vitro* to the constituent p15 and p12.

Attempts to purify the protease to homogeneity by conventional methods as described were unsuccessful. Although the protease could be concentrated considerably (in some cases 500-fold on protein basis), it was difficult to identify which protein was responsible for proteolytic activity. The protein patterns of concentrated phosphocellulose/Sephadex G-75 fractions made visible by staining after NaDodSO$_4$/PAGE (see Fig. 1) were very complex. More than 10 proteins were detected in the 10- to 20-kilodalton region of NaDodSO$_4$/PAGE, where the protease itself migrates (4).

To purify the protease in sufficient amounts and purity for structure analysis we utilized RP-HPLC. For these studies we first prepared an acetone powder from purified virus. From this powder we solubilized the protease by stepwise extraction with PD buffer having increasing concentrations

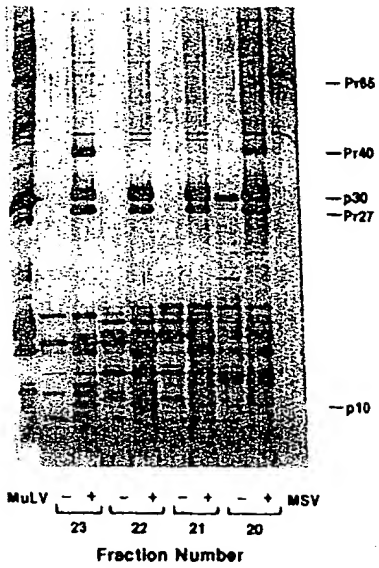


FIG. 1. Cleavage of Gz-MSV Pr65^{gag} with partially purified protease of Mo-MuLV. Purified disrupted virus (50 mg) was fractionated by a combination of phosphocellulose and Sephadex G-75 chromatography, and the protease activity of each fraction was determined as described in the text. Protein composition and protease activity of Sephadex G-75 fractions 20–23 are shown: −, fraction alone; +, fraction plus Gz-MSV Pr65^{gag}. The gel was stained with silver nitrate.

of NaCl. Much of the protease activity was extracted with 0–0.5 M NaCl in PD buffer, while most of the membrane proteins, including p15E and p15, stayed in the insoluble residue. The protease-active extracts were pooled, lyophilized, and dissolved in 3 ml of PD buffer, then fractionated on a Sephacryl S200 column in the cold and further purified by RP-HPLC using a μBondapak C_{18} column as shown in Fig. 2*A*. Fractions (5 ml) were collected and lyophilized to recover proteins. Purity and protease activity were determined by NaDodSO$_4$/PAGE analysis as shown in Fig. 2 *B* and *C*, respectively. The peak activity was eluted at about 33% acetonitrile (fraction 24 of Fig. 2) and clearly separated from p30, p12, p10, and other low molecular weight proteins. The purified protein showed a single band in NaDodSO$_4$/PAGE (Fig. 2*B*). When incubated with disrupted Gz-MSV it cleaved Pr65^{gag} to produce Pr40^{gag}, Pr27^{gag}, and p30 (Fig. 2*C*). The total protein recovered in RP-HPLC fractions 22–24 was 14 μg.

In the absence of a quantitative assay for the protease the determination of the recovery of enzymatic activity is difficult. If we define a unit as the activity (per unit volume) capable of 50% reduction of Pr65^{gag} band intensity after 16-hr incubation (see *Materials and Methods*), we can estimate that we extracted a total of 110 units of activity from the virus and found 66 units in HPLC fractions (Fig. 2). This corresponds to 60% overall recovery of protease activity. The possibility for the actual protease being a minor component copurifying with the protein peak cannot be completely excluded. However, this is unlikely since in NIH/3T3 cells transfected with cloned viral DNA having deletions only in the protease region, Pr65^{gag} is synthesized but not processed into mature protein components (unpublished observations, and S. Crawford and S. P. Goff, personal communication).

NH$_2$-terminal Amino Acid Sequence of the Protease. To determine the NH$_2$-terminal amino acid sequence of purified protease recovered from fraction 24, 0.5 nmol of protein was degraded in a single microsequence analysis. The amino

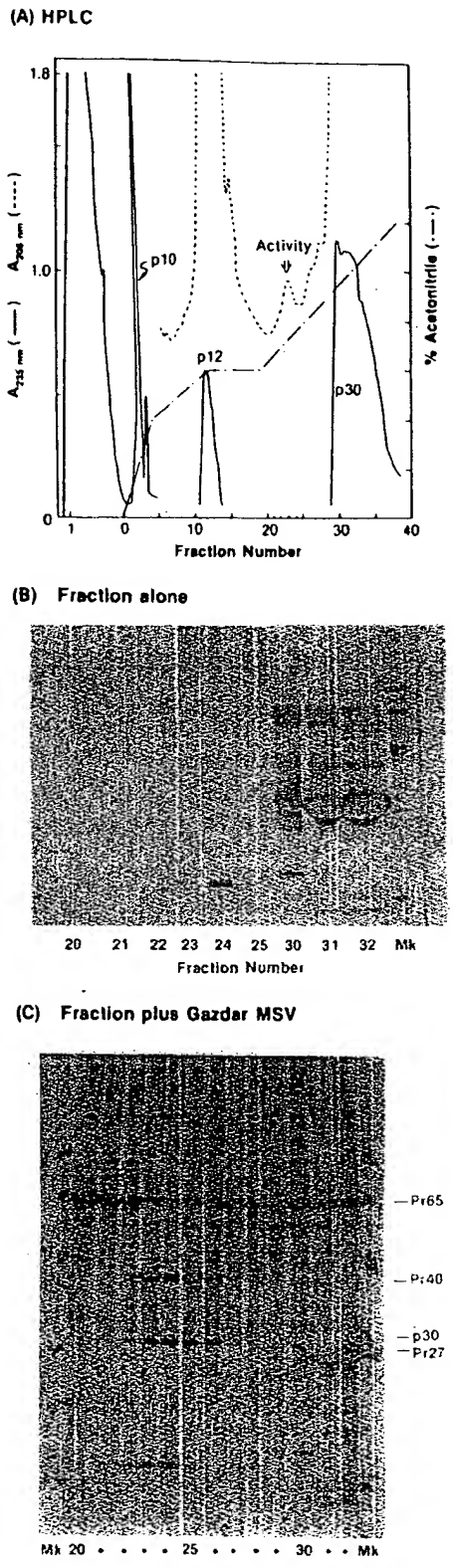


FIG. 2. Purification of protease by RP-HPLC. Purified virus (50 mg) was extracted with acetone. Protease activity extracted from acetone powder was further fractionated by Sephacryl S200 and then by RP-HPLC (μBondapak C_{18} column 0.39 × 30 cm, Waters Associates). (*A*) Absorbance profile. Gradient conditions were 0–20% (vol/vol) acetonitrile over 20 min; 20–30% acetonitrile over 40 min; isocratic at 30% acetonitrile for 30 min, and 30–60% acetonitrile over 90 min at a constant flow rate of 1.0 ml/min. (*B*) Purity of RP-HPLC-separated proteins by NaDodSO$_4$/PAGE. One-twentieth of each fraction was lyophilized and analyzed; staining was with silver nitrate. Lane Mk contained molecular weight markers: phosphorylase b, 92,000; bovine serum albumin, 68,000; ovalbumin, 46,000; carbonic anhydrase, 29,000; lysozyme, 14,400. (*C*) Assay of RP-HPLC fractions for protease activity. One-twentieth of each fraction was lyophilized and assayed for protease activity as described in text. Proteins are visualized by silver staining. Lanes Mk as in *B*.

```
                    5                                  10
Thr-Leu-Asp-Asp-Gln-Gly-Gly-Gln- X -Gln-
240 575 300 490 240 140 220 230     195

                   15                                  20
Glu-Pro-Pro-Pro-Glu- X -Arg-Ile-Thr-Leu-
180  65  90  85 110     135  50  50  70
```

FIG. 3. NH_2-terminal sequence of Mo-MuLV protease. The number under each assigned residue is the recovery (in pmol) of that residue. X, an unidentified residue.

acids identified at each cycle (first 20) are shown in Fig. 3 together with the quantitative yields.

COOH-Terminal Sequence Analysis. Purified protease was digested for various time periods with carboxypeptidase Y and the amino acids released were determined on the analyzer. The data shown in Table 1 allow us to conclude that the COOH terminus of the protease is leucine. Other amino acids released in smaller quantities were valine, glutamine (or serine), and proline. The kinetic analysis data do not by themselves define an accurate sequence, but they could be interpreted by comparison with the amino acid sequence deduced from the DNA sequence (12) as will be discussed below.

Position of Protease on Viral Genome and Suppression of Amber Codon into Glutamine. To determine whether the protease protein is virus encoded or not, we aligned the NH_2- and COOH-terminal amino acid sequences with nucleotide sequences. As shown in Fig. 4, the protease amino acid sequence starts with threonine encoded by triplet 2223–2225 and includes the last four amino acids of the *gag* region. Furthermore, the amber codon (UAG) is translated as glutamine, which is residue five of the protease. This is followed by a glycine residue encoded by the first triplet of the *pol* gene, indicating that translation continues in the same reading frame. These results clearly show that the $Pr65^{gag}$-specific protease is a virus-encoded enzyme and that it is synthesized by reading through the termination codon.

It is also seen that the determined NH_2-terminal sequence for the protease (15 residues are shown) matches the amino acid sequence predicted from the nucleotide sequence of proviral DNA designated pMLV-1 (12) except at position 11, where the protein has Glu instead of Asp. At this juncture it is important to point out that the pMLV-1 clone is not infectious and that the infectious clone (pMLV-48) of Miller and Verma (22), like the protein, also has Glu in position 11. A single base change, C → G, in the codon accounts for this difference. A comparison of the COOH-terminal sequence analysis results with the translated sequence of proviral DNA indicates that the COOH-terminal Leu must be the codon 2595–2597, which is adjacent to the NH_2-terminal Thr of reverse transcriptase (11). This determines that the COOH-terminal sequence is Pro-Leu-Gln-Val-Leu-OH and that the protease is composed of 125 amino acids (Figs. 4 and 5).

DISCUSSION

We have succeeded in purifying from Mo-MuLV a virus-encoded protease that is capable of processing *in vitro* the *gag* precursor polyprotein $Pr65^{gag}$ into the constituent struc-

Table 1. COOH-terminal amino acid sequence analysis of Mo-MuLV protease

Digestion time, min	Amino acids released,* pmol			
	Leu	Val	Gln	Pro
0.5	70	29	0	0
1.0	101	30	33	0
2.0	123	31	45	30
5.0	140	48	49	31
10.0	158	67	55	48
20.0	191	72	62	60

*Each sample analyzed at each time point had 200 pmol of protein.

tural proteins, which appear to be the same as those occurring in mature virions. We have definitely identified cleavage products p30, p10, and p15. The identification of p15 was aided by the high sensitivity of detection after its specific labeling with [^{3}H]myristate (unpublished data). The fourth *gag*-gene encoded protein, p12, however, could not be identified with absolute certainty in our assay system due to its poor affinity for Coomassie blue and to the apparent lack of specificity of the more sensitive silver stain (nucleic acids are stained equally well as proteins).

The NH_2- and COOH-terminal sequences of the Mo-MuLV protease as determined in this study and the availability of DNA sequences now present an opportunity for deducing the complete primary structure of this virus-encoded proteolytic enzyme. The complete amino acid sequence of the Mo-MuLV protease aligns (Fig. 5) without gaps with the amino acid sequences of putative proteases as inferred from DNA sequences of AKR mouse leukemia virus (AKV) (23), feline leukemia virus (FeLV) (24), and baboon endogenous virus (BaEV) (25). It is seen from this alignment that there are only four amino acid differences between the two mouse proteases compared. With respect to the Mo-MuLV sequence, FeLV and BaEV proteases have 25 (20%) and 39 (31.2%) changes, respectively, indicating highly conserved primary structures. From the combined results it appears that there are three variable regions: one at the NH_2-terminal region (residues 1–12), the second in the middle part of the molecule (residues 60–77), and the third at the COOH-terminal region (residues 108–120). The last five residues of the MuLV and FeLV sequences are identical, while in the BaEV sequence there are two substitutions, Leu → Ile and Val → Ile. These changes, however, would not significantly alter the nature of the cleavage site between the protease and reverse transcriptase. The regions involving residues 29–41 are identical among the sequences, and another highly conserved long stretch is present, extending from residue 78 to residue 101. Little is known about the active site of the retroviral protease, but inhibition studies done with avian protease (8) suggested that cysteine may be involved. Furthermore, it was also shown that the mouse protease is inhibited by tosyllysyl chloromethyl ketone (TLCK) (4). It is known that TLCK can inhibit thiol proteases or similar enzymes just as well as serine proteases, and on occasions it has been effectively used to identify cysteines at the active site (26). Interestingly, with the exception of BaEV, each of the viral proteases, including those of avian myeloblastosis virus and avian sarcoma virus, has only a single cysteine (residue 88 in the alignment of Fig. 5), which is preceded by Asp or Glu. It will be important to develop quantitative assays, perhaps utilizing synthetic peptide substrates (27), for the retroviral proteases to characterize them more completely.

Our results, together with the previously determined NH_2- and COOH-terminal sequences of reverse transcriptase (11) and their alignments with nucleotide sequences, suggest the map order for $Pr180^{gag\text{-}pol}$ of the mouse retrovirus to be 5′-p15–p12–p30–p10–protease–reverse transcriptase–endonuclease-3′. The polyprotein itself most likely has no proteolytic activity. It remains to be seen how the active protease is generated. Autocatalysis or an initial cleavage by another enzyme (probably cellular) may be responsible.

The most significant result reported in this study relevant to virus replication is the finding that *in vivo* translation of the *pol* gene resulting in the synthesis of the precursor polyprotein $Pr180^{gag\text{-}pol}$ occurs through in-frame readthrough of the amber termination codon. While we do not know the exact mechanism by which glutamine is inserted at the termination site in the translation process taking place in mouse fibroblasts, we can assume that this insertion is accomplished via the misreading of the UAG codon by normal $tRNA^{Gln}$ due to the wobble in the 3′ position of the anticodon. A more remote possibility is suppression by a specific nonsense suppressor tRNA. Such tRNAs have been identified not only in prokaryotes but also recently in eukaryotes (28, 29). Suppression of termination codons has been proposed to occur in plant viruses (30–32) and alphaviruses (33, 34). It appears that plant and animal viruses are capable of effectively utilizing the translational readthrough mechanism to produce from a single initiation site different amounts of proteins and polyproteins required for specific functions. The importance of this translational control for virus replication, infectivity, and pathogenicity could be directly tested by utilizing mutants in which the respective termination codons are eliminated.

In Mo-MuLV the *gag* and *pol* genes are in the same reading frame. However, available nucleic acid sequences indicate that apparently this is not true for Rous sarcoma virus (14), FeLV (24), human T-cell leukemia virus (35), and bovine leukemia virus (36). As with the Mo-MuLV study, protein sequencing will reveal whether in-frame suppres-

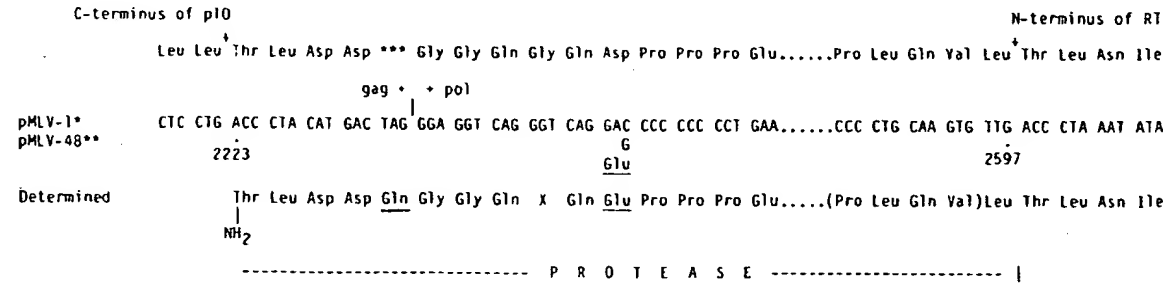


FIG. 4. Alignment of NH_2- and COOH-terminal amino acid sequences with DNA sequences of pMLV-1 (12) and pMLV-48 (21). The amber codon UAG is translated as glutamine (double underline). RT, reverse transcriptase.

FIG. 5. Amino acid sequence alignment of Mo-MuLV protease with corresponding sequences of AKR mouse leukemia virus (AKV) (23), feline leukemia virus (FeLV) (24), and baboon endogenous virus (BaEV) (25) inferred from nucleotide sequences. Corresponding to position 5 (Gln in Mo-MuLV and X in the other sequences) the genes have an amber termination codon. In this alignment the predicted NH_2-terminal sequence (positions 1–12) of the putative FeLV protease is translated from the nucleotide sequence of Laprevotte *et al.* (24) in the *gag* reading frame. The remainder of the sequence, starting with Pro in position 12, is in a reading frame as published (24). The DNA sequence between codon GAC (Asp in position 11) and CCC (Pro in position 12) has an additional C which was not decoded for the purposes of this alignment. It remains to be seen whether the FeLV suppression occurs in frame, as in MuLV, or requires a frameshift. The third possibility of course is splicing, as proposed by Laprevotte *et al.* (24).

sion, frameshift suppression, or splicing is involved in these latter cases. In all cases the elucidation of actual mechanism responsible for suppression will require further biochemical experiments with purified tRNAs.

We thank Cathy Hixson and Young Kim for their excellent technical assistance and Jeannie Clarke for preparation of the manuscript. This research was sponsored in part by the National Cancer Institute under contract no. N01-C0-23909 with Litton Bionetics, Inc.

1. Dickson, C., Eisenman, R., Fan, H., Hunter, E. & Teich, N. (1982) in *Molecular Biology of Tumor Viruses*, eds. Weiss, R. A., Teich, N. M., Varmus, H. E. & Coffin, J. M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 513–648.
2. Jamjoom, G. A., Naso, R. B. & Arlinghaus, R. B. (1977) *Virology* **78**, 11–34.
3. Philipson, L. P., Andersson, V., Olshevsky, R., Weinberg, D., Baltimore, D. & Gestland, R. (1978) *Cell* **13**, 189–199.
4. Yoshinaka, Y. & Luftig, R. B. (1977) *Cell* **12**, 709–719.
5. Witte, O. N. & Baltimore, D. (1978) *J. Virol.* **26**, 750–761.
6. Yoshinaka, Y. & Luftig, R. B. (1982) *Virology* **118**, 380–388.
7. Von der Helm, K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 911–915.
8. Dittmar, K. J. & Moelling, K. (1978) *J. Virol.* **28**, 106–118.
9. Traktman, P. & Baltimore, D. (1980) in *Biosynthesis, Modification, and Processing of Cellular and Viral Proteins*, eds. Koch, G. & Richter, D. (Academic, New York), pp. 301–306.
10. Levin, J. G., Hu, S. C., Rein, A., Messer, L. I. & Gerwin, B. (1984) *J. Virol.* **51**, 470–478.
11. Copeland, T. D., Gerard, G. F., Hixson, C. G. & Oroszlan, S. (1985) *Virology*, in press.
12. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–548.
13. Sauer, R. T., Allen, D. D. W. & Niall, H. D. (1981) *Biochemistry* **20**, 3784–3791.
14. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32**, 853–869.
15. Gazdar, A. F., Phillips, L. A., Sarma, P. S., Peebles, P. T. & Chopra, H. C. (1971) *Nature (New Biol.) (London)* **234**, 69–72.
16. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
17. Wray, W., Boulikas, T., Wray, V. P. & Hancock, R. (1981) *Anal. Biochem.* **118**, 197–203.
18. Henderson, L. E., Sowder, R., Copeland, T. D., Smythers, G. & Oroszlan, S. (1984) *J. Virol.* **52**, 492–500.
19. Copeland, T. D., Grandgenett, D. P. & Oroszlan, S. (1980) *J. Virol.* **36**, 115–119.
20. Henderson, L. E., Copeland, T. D. & Oroszlan, S. (1980) *Anal. Biochem.* **102**, 1–7.
21. Hayashi, R. (1976) *Methods Enzymol.* **45**, 568–587.
22. Miller, A. D. & Verma, I. M. (1984) *J. Virol.* **49**, 214–222.
23. Herr, W. (1984) *J. Virol.* **49**, 471–478.
24. Laprevotte, I., Hampe, A., Sherr, C. J. & Galibert, F. (1984) *J. Virol.* **50**, 884–894.
25. Tamura, T. (1983) *J. Virol.* **47**, 137–145.
26. Whitaker, J. R. & Perez-Villasenor, J. (1968) *Arch. Biochem. Biophys.* **124**, 70–76.
27. Copeland, T. D. & Oroszlan, S. (1982) in *Peptides: Synthesis, Structure, Function*, eds. Rich, D. H. & Gross, E. (Pierce, Rockford, IL), pp. 497–500.
28. Diamond, A., Dudock, B. & Hatfield, D. (1981) *Cell* **25**, 497–506.
29. Hatfield, D. L., Dudock, B. S. & Eden, F. C. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4940–4944.
30. Pelham, H. R. B. (1978) *Nature (London)* **272**, 469–471.
31. Goelet, P., Lomossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. R. & Karn, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5818–5822.
32. Morch, M. D., Drugon, G. & Benicourt, C. (1982) *Virology* **119**, 193–198.
33. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5271–5275.
34. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984) *Virology* **133**, 92–110.
35. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3618–3622.
36. Rice, N. R., Stephens, R. M., Burny, A. & Gilden, R. V. (1985) *Virology*, in press.